NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

NDVI AND PANCHROMATIC IMAGE CORRELATION USING TEXTURE ANALYSIS	
by	
David A. Jablonski	
March 2010	
Thesis Advisor:	Richard C. Olsen
Second Reader:	David M. Trask

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2010	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE NDVI and Panchromatic Image Correlation Using Texture Analysis			**5. FUNDING NUMBERS**
6. AUTHOR(S) David A. Jablonski			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number: _____________.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

The purpose of this research is to apply panchromatic satellite imagery to the task of locating kelp in the California coastal waters. The task is currently done using multi-spectral imagery (MSI), but there are time intervals wherein only panchromatic data are available. Panchromatic images were analyzed using various threshold approaches, analysis techniques, and texture analysis. Results were then compared to MSI data analyzed using the standard Normalized Difference Vegetation Index (NDVI). Four classification methods were used: Maximum Likelihood, Mahalanobis Distance, Minimum Distance, and Binary Encoding. The main problem with this approach was sunglint off of the water. It proved difficult to eliminate all of it in the classification of kelp. The Receiver Operating Characteristic (ROC) curves proved that the panchromatic and variance texture feature images were well above the line of no-discrimination, so they are a very good detector and discriminator of kelp and water. Using panchromatic and variance in the Mahalanobis Distance, and Minimum Distance classification methods, the result is an overall accuracy of 98.5% of the Santa Barbara Coastal Long-Term Ecological Research (SBC-LTER) Program research areas of Arroyo Burro and Mohawk.

14. SUBJECT TERMS NDVI, Panchromatic, Texture analysis, Kelp Detection			**15. NUMBER OF PAGES** 89
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

NDVI AND PANCHROMATIC IMAGE CORRELATION USING TEXTURE ANALYSIS

David A. Jablonski
Captain, United States Air Force
B.S., Missouri University S&T, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SPACE SYSTEMS OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
March 2010

Author:	David A. Jablonski
Approved by:	Richard C. Olsen Thesis Advisor
	David M. Trask Second Reader
	Rudolf Panholzer Chair, Space Systems Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The purpose of this research is to apply panchromatic satellite imagery to the task of locating kelp in the California coastal waters. The task is currently done using multi-spectral imagery (MSI), but there are time intervals wherein only panchromatic data are available. Panchromatic images were analyzed using various threshold approaches, analysis techniques, and texture analysis. Results were then compared to MSI data analyzed using the standard Normalized Difference Vegetation Index (NDVI). Four classification methods were used: Maximum Likelihood, Mahalanobis Distance, Minimum Distance, and Binary Encoding. The main problem with this approach was sunglint off of the water. It proved difficult to eliminate all of it in the classification of kelp. The Receiver Operating Characteristic (ROC) curves proved that the panchromatic and variance texture feature images were well above the line of no-discrimination, so they are a very good detector and discriminator of kelp and water. Using panchromatic and variance in the Mahalanobis Distance, and Minimum Distance classification methods, the result is an overall accuracy of 98.5% of the Santa Barbara Coastal Long-Term Ecological Research (SBC-LTER) Program research areas of Arroyo Burro and Mohawk.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF EQUATIONS

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I would like to express my thanks to my wife, Priscilla, for allowing the time away to complete this thesis.

Next, I would like to thank my parents, Joseph and Leona Jablonski, for their love and support over the course of my life. This thesis would not have been possible without all their guidance throughout the years.

I’d like to thank Richard Olsen for taking on a distance learning student and helping me in all aspects of this thesis. I’d also like to thank Angie and Krista for their help, especially for the weekends when you all did not have to be there.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The health of the coastal environment in California depends intimately on the health of the kelp forests in the near coastal area. This aquatic environment is home to one of the most diverse ecosystems on the planet. A kelp bed is a highly dynamic ecosystem and can vary in size over days, weeks, and months. Kelp forests provide a habitat for marine organisms and are a source for understanding many ecological processes. They have been the focus of extensive research and continue to provide important ideas that are relevant beyond this ecosystem. For example, kelp forests can influence coastal oceanographic patterns (Jackson, 1983). The influence of humans has often contributed to kelp forest degradation. Of particular concern are the effects of overfishing near shore ecosystems, which can release herbivores from their normal population regulation and result in the over-grazing of kelp and other algae (Sala, 1998). The implementation of marine protected areas (MPAs) is one management strategy useful for addressing such issues, since it may limit the impacts of fishing and buffer the ecosystem from additive effects of other environmental stressors. The Santa Barbara Coastline is now federally protected, and researchers map its growth to track the health of the ecosystem.

This is why it is important to measure the health and extent of the kelp forest, and monitor changes in this ecosystem. This can be done using satellite imagery, in particular multi-spectral imagery (MSI). The University of California, Santa Barbara, Institute for Computational Earth System Science (UCSB ICESS) is working on a project that could benefit from this sort of image processing. The institute's research project involves mapping the size of the kelp bed in the Santa Barbara Channel. "The goal of the Santa Barbara Coastal Long-Term Ecological Research (SBC-LTER) is to evaluate whether land use patterns in local watershed influence kelp forest ecosystems through the run-off of nutrients (fertilizers), sediments, and other pollutants" (Lenihan, 2004). Short time periods between image acquisitions can help give a more accurate picture of the kelp bed.

The advent of high-spatial resolution civil imaging systems includes sensors that only provide panchromatic imagery. ICESS would like to use images with better resolution to map the kelp bed with greater accuracy. It is the purpose of this thesis to study the utility of such imagery for the purpose of monitoring kelp forests.

This thesis will look to further explore the different systems and processing techniques that will be used for this research. The limitations will be discussed, as well as possible areas of improvement. Sunglint or other bright objects on the water's surface create confusion in the results for the panchromatic image. In the second part of this thesis, the goal is to prove that the high spatial resolution of the panchromatic data from the QuickBird sensor can be used to mitigate these errors.

II. BACKGROUND

A. UCSB KELP PROJECT

The Giant Kelp forest bed is a very large and important ecosystem. The Santa Barbara Coastal Long-Term Ecological Research (SBC-LTER) Program, funded by the National Science Foundation, was founded to study this “long-term ecological phenomena” (Lenihan, 2004). “The goal of the SBC-LTER is to evaluate whether land use patterns in local watershed influence kelp forest ecosystems through the run-off of nutrients (fertilizers), sediments, and other pollutants” (Lenihan, 2004). There are several research objectives that must be accomplished to achieve this goal. The first objective is to examine “how nutrient inputs from the land and ocean influence the standing crop and production of giant kelp” (Lenihan, 2004). The next objective is to take biomass data acquired by the kelp harvesting industry from as far back as 1958 along the southern California coast to analyze historical trends. Another research objective of the SBC-LTER team is to work with oceanographers “to determine how nutrients and sediments are transported and where they end up, and the ecological effects of these inputs to the kelp forest” (Lenihan, 2004). The last objective of the program is described in the following paragraph.

The major research objective that pertains to satellite imagery is the measurement of giant kelp, pictured in Figures 1 and 2, which has the scientific name of *Macrocystis pyrifera*. There are two types of measurements that the team is trying to collect. The first measurement is that of the canopy cover. Just like a tree’s canopy, kelp’s canopy is that which is seen from the surface of the water. Normal pictures and observations can be used to calculate this data. The next and more important measurement is that of the biomass data. There are a few ways to measure the amount of biomass of kelp in the water. The first is to have divers in the water, physically measuring the kelp. Not only is this method time consuming, but it is also costly. This is where satellite imagery comes into play. One satellite image can cover the entire area of the kelp beds. Then, using processing techniques such as Normalized Difference Vegetation Index and texture feature analysis, this biomass information can be calculated much faster and with less effort. To better understand the subject of this research, the characteristics of kelp will be further defined.

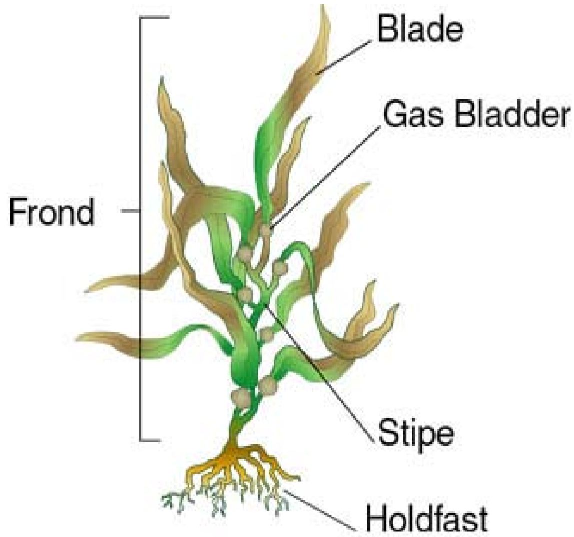


Figure 1. *Macrocystis pyrifera*, or giant kelp (From Cavanaugh, 2008).

Figure 2. Kelp bed (Image available at University of California Natural Reserve System, http://nrs.ucop.edu/SP10-Santa-Barbara-Ecology.htm)

Kelp forests are underwater areas with a high density of kelp. They are recognized as one of the most productive and dynamic ecosystems on earth. Kelp is known as the "ecosystem engineer" because it provides the physical substrate and habitat about which kelp forest communities are built (Jones, 1997). Kelp is defined by three basic structural units: the holdfast, stipe, and frond (Dayton, 1985). The holdfast is a root-like mass that anchors the kelp to the sea floor. Unlike true roots, however, it is not responsible for absorbing and delivering nutrients to the rest of the plant. The stipe is like a plant stalk, extending up from the holdfast and providing a support framework for other morphological features. The fronds are leaf- or blade-like attachments extending from the stipe, sometimes along its full length, and are the sites of nutrient uptake and photosynthetic activity. Many kelp species have gas-filled bladders usually located at the base of fronds near the stipe. These structures provide the necessary buoyancy for kelp to maintain an upright position in the water column. All of these features can be seen in Figures 1 and 2. The life cycle of kelp is highly dynamic one. Kelp has an average frond life of 3–5 months and an average plant life of 2–3 years. Kelp has growth rates up to 0.5m per day, which explains the need to get measurements weekly or at least monthly. As seen in Figure 3, kelp beds can vary in size over a matter of months (Cavanaugh, 2008).

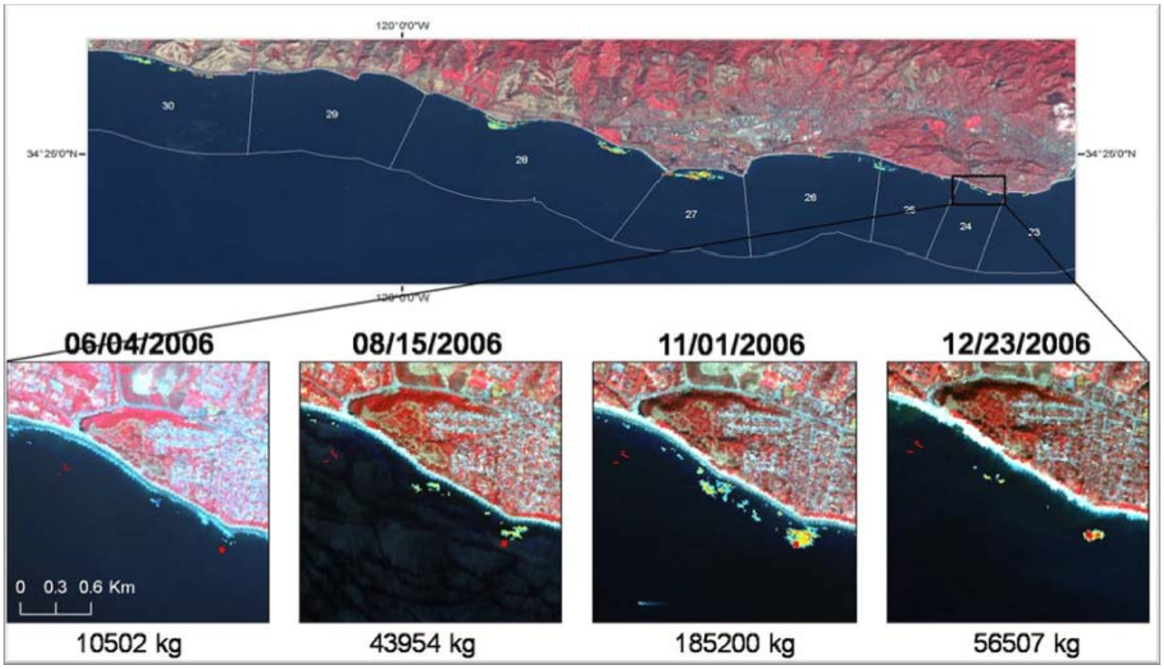


Figure 3. Kelp growth over six months (From Cavanaugh, 2008)

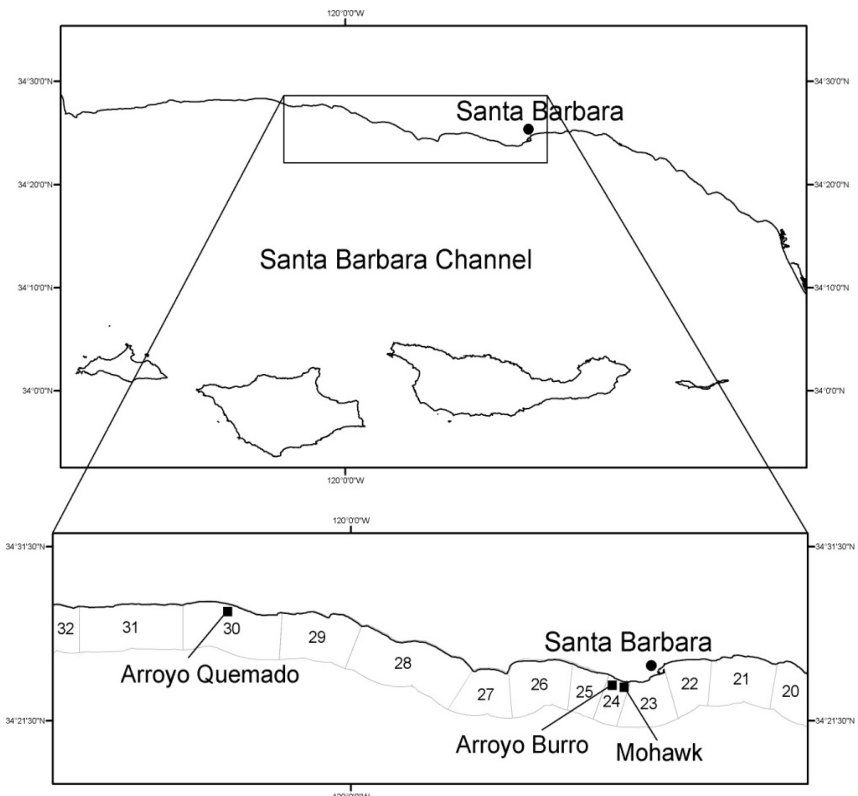


Figure 4. SBC-LTER area (From Cavanaugh, 2008)

The large image for the thesis ranges between areas 21 and 26 in Figure 4. The study area image includes the areas of Arroyo Burro and Mohawk, but not the Arroyo Quemado area.

B. LAND VERSUS WATER VEGATATION MEARSUREMENT

There is a difference between analyzing images of vegetation on land as compared to an image of vegetation in water. On land, there are many different objects from which light can be reflected. This can be anything from a truck to an animal, so locating the vegetation in the panchromatic image purely on an intensity scale would be more difficult. The situation is completely different when analyzing plants in water. Water is a black body that soaks up all of the incoming light. The only thing to reflect the light is the vegetation along with a few outliers. These outliers can include boats on the water and sunglint, which will be discussed later.

C. SPECTRAL CHARACTERISTICS OF VEGETATION

The study will focus on classifying kelp. Kelp uses photosynthesis to convert carbon dioxide and water to sugar. Plants use a green pigment, chlorophyll, to implement the chemical conversion. It is chlorophyll that is responsible for the predominate spectral signature of kelp and other living plants. The unique spectral response of live vegetation is the high near-infrared reflectance coupled with a low red reflectance. Processing the near IR and Red channels of a multispectral image using NDVI enables the researcher to determine and locate vegetation. This processing algorithm does not work for panchromatic imagery because there is only one channel, usually covering the 0.4 to 0.8 um region. As seen in Figure 5, the high near-infrared dominates the spectral response. However, green grass (chlorophyll rich) is a very bright object even in the panchromatic image. This should help the classification methods to be able to classify kelp.

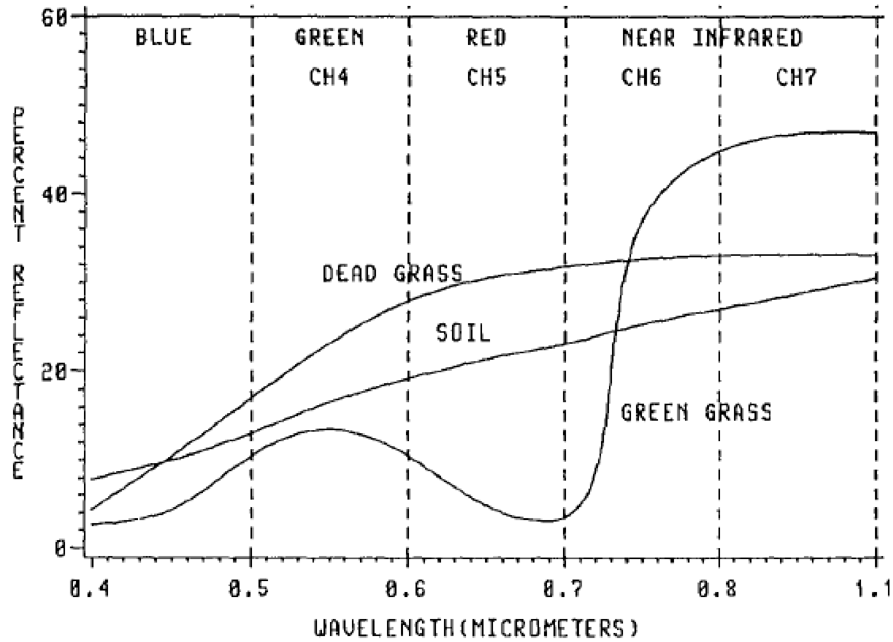


Figure 5. Spectral reflectance of vegetation and soil from 0.4 to 1.1 mm (From Perry & Lautenschlager, 1984)

D. TEXTURE THEORY

1. Texture Features for Image Classification (Haralick, 1973)

One powerful tool for extracting information from panchromatic imagery is texture analysis. The intrinsically higher spatial resolution of the panchromatic satellite imagers provides superb texture information that can be exploited in an attempt to compensate for the loss of multi-spectral information.

The 1973 study by Haralick has become the basis of reference in many image classification studies since its publication. It laid the foundation of how to examine Grey-Tone images. The texture features derived in the Haralick study will be used to classify the images into kelp or water. It is important to understand the derivation of these texture features so that results of the classification can be understood. First, the set up of the Gray-Tone Spatial-Dependence Matrices will be discussed, and then the equations for each texture feature. This study does not go into a detailed description of texture features.

Satellite imagery is generally stored in a computer as a two-dimensional array. The X spatial domain is $L_x = \{1, 2, \ldots, N_x,\}$, or samples. The N_x value for the X domain will generally be in the thousands. The Y spatial domain is $L_y = \{1,2,\ldots,N_y$, or lines.} The $L_x \times L_y$ is the set of resolution cells and must have some gray-tone value $G = \{1,2,\ldots,N_g\}$ to each and every resolution cell. Current satellite systems generally have a dynamic range of 11–12 bits, for a typical dynamic range of 0–2047 or 0–4095. The texture techniques described below generally require a reduced dynamic range of 0–63 or 0–255 due to processing limitations. Several different types of image processing tasks such as coding, restoration, enhancement, and classification can be performed with the information just stated.

The Gray-Tone Spatial-Dependence Matrices are described by Lx, Ly, and G. There are four closely related measures that are termed angular nearest-neighbor gray-tone spatial-dependence matrices. These matrices are the ones by which all of the texture features will be calculated. Figure 6 shows a 3×3 matrix, which is the size used in this thesis at the angles of 0, 45, 90, and 135.

6	7	8
5	*	1
4	3	2

Figure 6. Nearest Neighbor set up, resolution cells 1 and 5 are 0 degrees nearest neighbors to resolution cell *; resolution cells 2 and 6 are 135 degrees nearest neighbors; resolution cells 3 and 7 are 90 degrees nearest neighbors; and resolution cells 4 and 8 are 45 degrees nearest neighbors to *. (From Haralick, 1973)

$$
\begin{aligned}
P(i,j,d,0^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
\times (L_y \times L_x) \mid k - m = 0, |l - n| = d, \\
I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,45^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
\times (L_y \times L_x) \mid (k - m = d, l - n = -d) \\
\text{or } (k - m = -d, l - n = d), \\
I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,90^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
\times (L_y \times L_x) \mid |k - m| = d, \\
l - n = 0, I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,135^\circ) = \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
\times (L_y \times L_x) \mid (k - m = d, l - n = d) \\
\text{or } (k - m = -d, l - n = -d), \\
I(k,l) = i, I(m,n) = j\} \qquad (1)
\end{aligned}
$$

where # denotes the number of elements in the set.

Equation 1. Nearest Neighbor Equations (From Haralick, 1973)

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

(a)

		Grey Tone			
		0	1	2	3
Grey Tone	0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
	1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
	2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
	3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(b)

0°

$$P_H = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

(c)

90°

$$P_V = \begin{pmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

(d)

135°

$$P_{LD} = \begin{pmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

(e)

45°

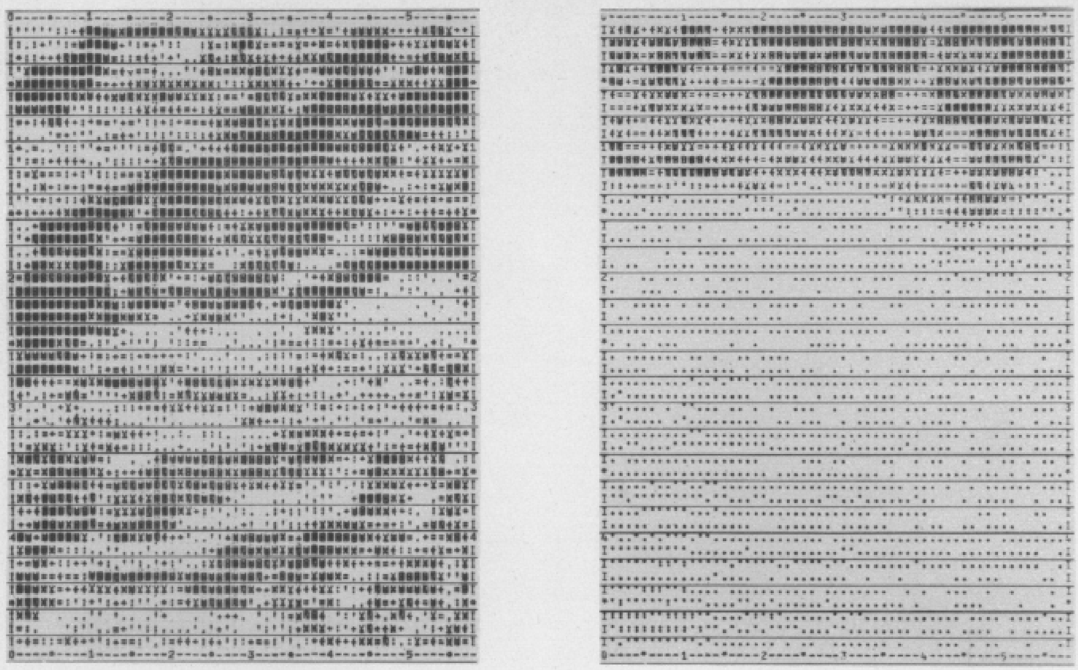
$$P_{RD} = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

(f)

(a) 4 × 4 image with four gray-tone values 0–3. (b) General form of any gray-tone spatial-dependence matrix for image with gray-tone values 0–3. #(i,j,) stands for number of times gray tones I and j have been neighbors. (c) – (f) Calculation of all four distance 1 gray-tone spatial-dependence matrices.

Figure 7. Nearest Neighbor Matrices (From Haralick, 1973)

Now, the Nearest Neighbor Matrices in Figure 7 are ready to have the textural information extracted out of them. The study uses three texture features as examples to describe the information that can be extracted and they are angular second-moment feature (ASM), contrast feature, and correlation feature. These are the same that will be used for this thesis and their use will be the same. The angular second-moment feature (ASM), which has been renamed as the homogeneity texture feature, is just that—the measure of homogeneity of the image. In a homogeneous image, such as shown in water body image in Figure 8, there are very few dominant gray-tone transitions. “The contrast feature is a difference moment of the P matrix and is a measure of the contrast or the amount of local variations present in an image" (Haralick, 1973). Since there is a lot of variation in the grassland image in Figure 8 as compared to the water body, the contrast feature for the grassland image has higher values compared to the water body image. "The correlation feature is a measure of gray-tone linear-dependencies in the image" (Haralick, 1973). For the grassland and water body images, the correlation texture feature is somewhat higher in the horizontal. In this thesis, only the average is computed for each texture feature. The water-body image mostly has a constant gray-tone value plus some noise which is the sunglint. Since the sunglint pixels are uncorrelated, the correlation texture feature has lower values for the water body as compared to the grassland image.



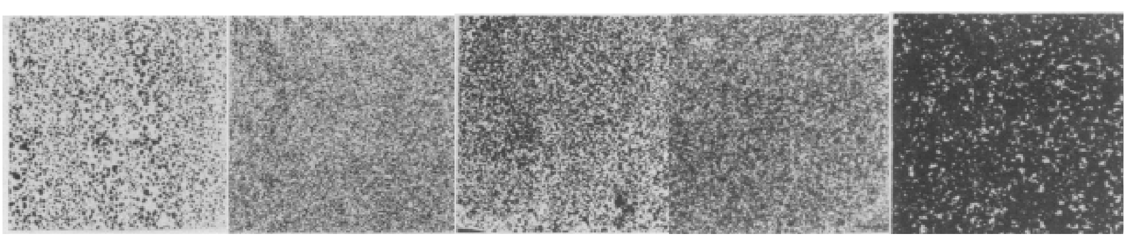
	Grassland			Water Body		
Angle	ASM	Contrast	Correlation	ASM	Contrast	Correlation
0°	.0128	3.048	.8075	.1016	2.153	.7254
45°	.0080	4.011	.6366	.0771	3.057	.4768
90°	.0077	4.014	.5987	.0762	3.113	.4646
135°	.0064	4.709	.4610	.0741	3.129	.4650
Avg.	.0087	3.945	.6259	.0822	2.863	.5327

Figure 8. Three textural features for two different land-use category images (From Haralick, 1973)

Now that the study explained what some of the texture features are and what they can do, the study moves to the application of the texture features. There were three sets of data used to analyze: Photomicrographs of Sandstones, Aerial Photographic Data Set, and Satellite Imagery. All of the data sets apply, but it is the satellite imagery that applies directly to this thesis. The two different classification algorithms used are the Piecewise Linear Discriminate Function Method and the Min-Max Decision Rule. The results from the Photomicrographs of Sandstones and Aerial Photographic Data Set can be found in Figures 9 and 10.

		ASSIGNED CATEGORY					Total
		Dexter-L	Dexter-H	St. Peter	Upper Muddy	Gaskel	
TRUE CATEGORY	Dexter-L	29	0	1	0	0	30
	Dexter-H	0	15	0	0	0	15
	St. Peter	2	0	22	4	0	28
	Upper Muddy	0	0	4	11	0	15
	Gaskel	0	0	0	0	12	12
	TOTAL	31	15	27	15	12	100

Number of samples in test set = 100; number of samples in training set = 143; overall accuracy of classification of test set = 89%.

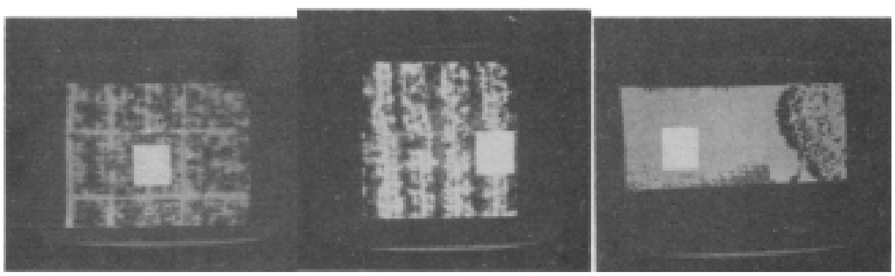


Dexter-L Dexter-H St. Peter Upper Muddy Gaskel

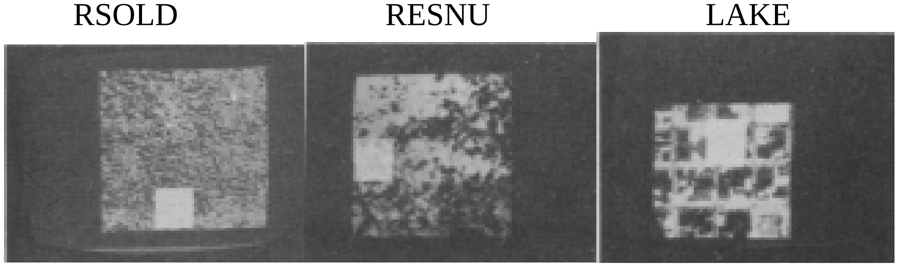
Figure 9. Accuracy Results for Classification of Photomicrographs of Sandstones (After Haralick, 1973)

		ASSIGNED CATEGORY								TOTAL
		RSOLD	RESNU	LAKE	SWAMP	MARSH	URBAN	RAIL	SCROD	
TRUE CATEGORY	RSOLD	17	0	0	0	1	1	0	1	20
	RESNU	1	15	0	0	1	2	1	0	20
	LAKE	0	0	19	0	0	0	0	1	20
	SWAMP	0	0	0	19	1	0	0	0	20
	MARSH	0	0	0	0	12	0	0	8	20
	URBAN	2	3	0	0	0	15	0	0	20
	RAIL	0	1	0	0	0	1	5	3	10
	SCROD	0	0	0	0	2	0	0	38	40
	TOTAL	20	19	19	19	17	19	6	51	170

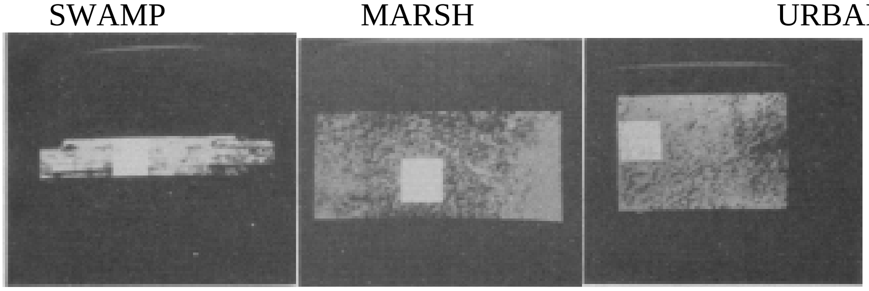
140 out of 170, or 82.3%, of the images were correctly classified



RSOLD RESNU LAKE



SWAMP MARSH URBAN



RAIL SCRUB WOOD (SCRUB)

Figure 10. Accuracy Results from the Classification of the Aerial Photographic Data Set (After Haralick, 1973)

The overall accuracy was 89% for the photomicrograph image set, 82% for the aerial photographs, and 83% for the satellite imagery. There are 14 equations, as described in Equation 2.

Notation

$p(i,j)$ — (i,j)th entry in a normalized gray-tone spatial-dependence matrix, $= P(i,j)/R$.

$p_x(i)$ — ith entry in the marginal-probability matrix obtained by summing the rows of $p(i,j)$, $= \sum_{j=1}^{N_g} P(i,j)$.

N_g — Number of distinct gray levels in the quantized image.

$\sum_i$ and $\sum_j$ $\quad \sum_{i=1}^{N_g}$ and $\sum_{j=1}^{N_g}$, respectively.

$$p_y(j) = \sum_{i=1}^{N_g} p(i,j).$$

$$p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{N_g} \sum_{j=1}^{N_g} p(i,j), \qquad k = 2,3,\cdots,2N_g.$$

$$p_{x-y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^{N_g} \sum_{j=1}^{N_g} p(i,j), \qquad k = 0,1,\cdots, N_g - 1.$$

Textural Features

1) Angular Second Moment:

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2.$$

2) Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}.$$

3) Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of p_x and p_y.

4) Sum of Squares: Variance

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i,j).$$

5) Inverse Difference Moment:

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i,j).$$

6) Sum Average:

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i).$$

7) Sum Variance:

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i).$$

8) Sum Entropy:[2]

$$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log \{p_{x+y}(i)\}.$$

9) Entropy:

$$f_9 = -\sum_i \sum_j p(i,j) \log (p(i,j)).$$

10) Difference Variance:

$$f_{10} = \text{variance of } p_{x-y}.$$

11) Difference Entropy:

$$f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log \{p_{x-y}(i)\}.$$

12), 13) Information Measures of Correlation:

$$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

$$f_{13} = (1 - \exp[-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = -\sum_i \sum_j p(i,j) \log (p(i,j))$$

where HX and HY are entropies of p_x and p_y, and

$$HXY1 = -\sum_i \sum_j p(i,j) \log \{ p_x(i) p_y(j) \}$$

$$HXY2 = -\sum_i \sum_j p_x(i) p_y(j) \log \{ p_x(i) p_y(j) \}.$$

14) Maximal Correlation Coefficient:

$$f_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}$$

where

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}.$$

Equation 2. Fourteen Equations of the set of 28 texture features (From Haralick, 1973)

2. Flood Hazard Assessment Using Panchromatic Satellite Imagery (Alhaddad, 2008)

A 2008 study by Alhaddad discusses the use of panchromatic satellite imagery for flood hazard mapping. This paper is relevant to this thesis because it shows how texture features can be used in classification methods. While it does not go into the specific texture features and how they affect the classification, it does describe four different classifications methods and how the results of each can differ depending on the terrain.

The study area was the Nile River in Egypt, and two SPOT images from 1997 and 1998 were used. The study used four different approaches that could be used for pan image classification and flood hazard assessment, image interpretation, edge detection, pixel-based image classification, and texture analysis.

This study looked at some of the previous work done in texture analysis and image classification like the Grey Level Co-occurrence Matrix (GLCM) by Haralick, 1973. First-order and second-order texture measures on GLCM consist of Standard Deviation, Range, Minimum, Maximum and Mean. The second order of texture measures includes Angular Second Moment, Contrast, Correlation, Dissimilarity, Entropy, Information Measures of Correlation, Inverse Difference Moment and Sum of Squares Variance. In Equation 3, the equations for Contrast, Dissimilarity, Mean, and Standard Deviation are shown.

$$Contrast = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}(i-j)^2 \qquad Mean = \mu = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} iP_{i,j}$$

$$Dis = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}|i-j| \qquad SD = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}(i-\mu)^2}$$

Where N = number of grey levels,
P = normalized symmetric GLCM of dimension N × N
Pij = is the (i,j)th element of P

Equation 3. The equations of Dissimilarity, Contrast, Mean and Standard Deviation (From Alhaddad, 2008)

For this study, three different land cover classes of Agricultural Land, Desert Area, and Water Bodies were used. Five supervised classification methodologies: Minimum Distance (MinD) and Maximum Likelihood (MLC), Artificial Neutral Network (ANN) Classifier, Contextual (CON) Classifier, and 5-Nearest Neighbor (kNN) Classifier were used to classify the terrain classes in the images. Five hundred random samples were classified by the five classifiers. Three rounds of classification were carried out using pan imagery only first and then texture only, then the combining of the two to compare the accuracy and the final computed flooding areas. This study goes into some of the processes and the time related to these four different approaches. Since only image classification is used in this thesis, this section is not applicable. The results of texture analysis are important to this thesis. Minimum Distance (MinD) and Maximum Likelihood (MLC) are the two used in this thesis and the results are shown in Figure 11. An interesting point not addressed in the study, but important to this thesis, is the difference between the different land terrains. The water was the most accurate followed by the desert and followed by the agricultural land. These go in order from the smoothest to the roughest and show that it is easier to more accurately classify a smooth object. The more noise introduced into an image, such as the sunglint, the harder it is to classify.

	SPOT 1997				SPOT 1998			
	Agricultural Land	Desert	Water Bodies	Overall Accuracy	Agricultural Land	Desert	Water Bodies	Overall Accuracy
PAN (ANN)	0.6278	0.9078	1.0000	86.8% (3)	0.9137	0.9562	0.9686	96.4% (1)
PAN (CON)	0.6430	0.9220	0.9090	86.2% (4)	0.7788	0.9552	0.9795	91.6% (4)
PAN (kNN)	0.5383	0.9734	1.0000	84.4% (5)	0.7901	0.9648	1.0000	93.4% (3)
PAN (MinD)	0.7526	0.9821	0.9794	92.4% (1)	0.9909	0.9646	0.7026	93.4% (3)
PAN (MLC)	0.7615	0.8922	0.9794	90.8% (2)	0.8217	0.9478	1.0000	94.0% (2)
TEX (ANN)	0.6592	0.9318	1.0000	88.4%(1)	0.7844	0.9731	1.0000	93.4% (1)
TEX (CON)	0.6027	0.9462	1.0000	85.6% (2)	0.7247	0.9901	0.9616	90.4% (3)
TEX (kNN)	0.6183	0.8547	1.0000	85.4% (3)	0.7337	0.9635	1.0000	91.6% (2)
TEX (MinD)	0.6546	0.6445	0.8766	80.6% (4)	0.7104	0.7598	0.7011	83.2% (5)
TEX (MLC)	0.5438	0.6663	1.0000	78.6% (5)	0.6695	0.9623	1.0000	89.4% (4)
PAN+TEX (ANN)	0.6585	0.9651	1.0000	89.0% (2)	0.7901	0.9646	1.0000	93.4% (2)
PAN+TEX (CON)	0.7152	0.9729	0.8297	88.8% (3)	0.7167	0.9641	0.9748	89.8% (5)
PAN+TEX (kNN)	0.6381	0.9734	1.0000	88.4% (4)	0.7709	0.9648	1.0000	92.8% (3)
PAN+TEX (MinD)	0.7243	0.9822	0.8954	90.8% (1)	0.8522	0.9646	0.8585	93.6% (1)
PAN+TEX (MLC)	0.6063	0.9078	1.0000	86.0% (5)	0.7217	0.9641	1.0000	91.2% (4)

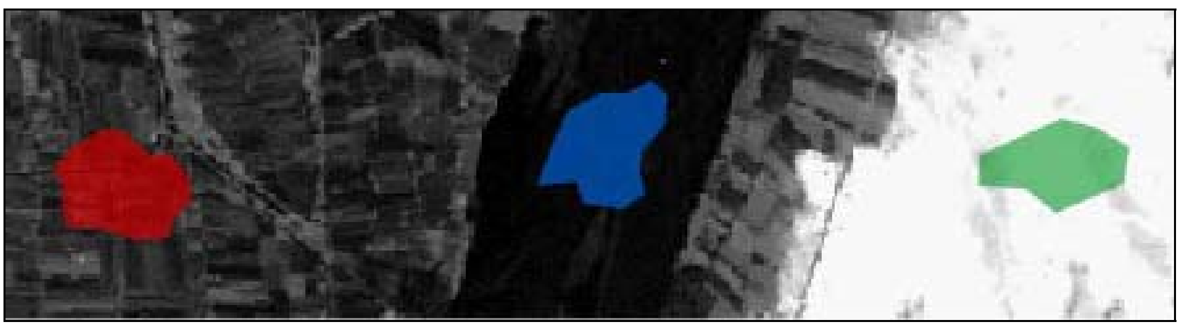


Figure 11. Accuracy results of the Different Approaches (After Alhaddad, 2008)

3. Study of Urban Spatial Patterns from SPOT Panchromatic Imagery Using Textural Analysis (From Shi, 2003)

A 2003 study by Shi discusses the use of texture features and how the addition of more texture features can help in accuracy. This paper is relevant to this thesis because it describes what each texture feature is doing to the image and how each texture feature relates to each other. This study helps to understand the how and why a certain texture feature would be used. This thesis uses both co-occurrence and occurrence texture features. The Shi study used the eight co-occurrence texture features and also included the Number of Different Grey-Levels (NDG) and Edge Density (ED). Both NDG and ED will not be further described since they do not apply. The images used for the Shi study are SPOT images of Beijing, China. The different terrain areas used for the study are shown in Figure 12.

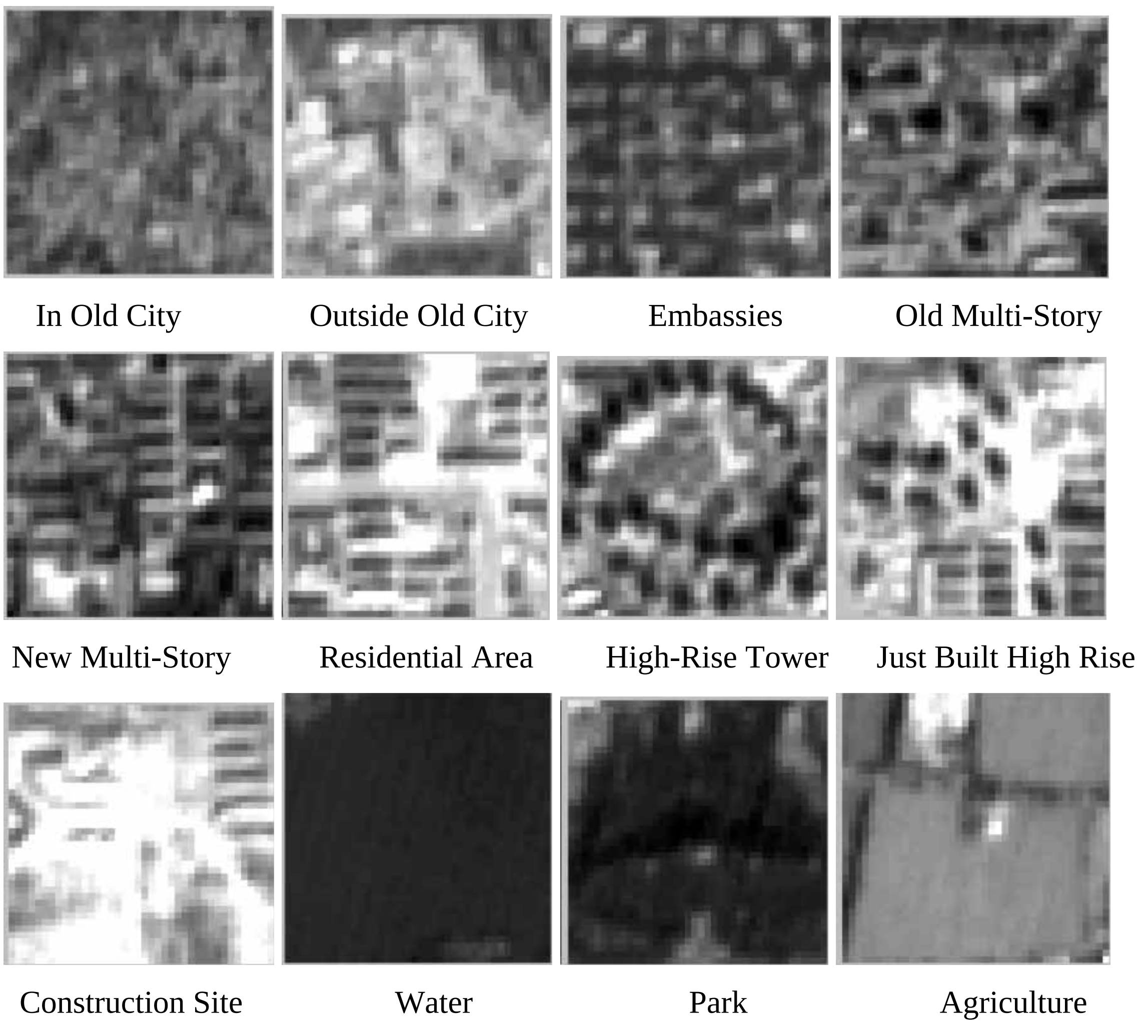


Figure 12. Samples of the different structures of the SPOT image (After Shi, 2003)

Eight texture features are homogeneity (HOM), contrast (CON), dissimilarity (DIS), mean (MEAN), standard deviation (SD), entropy (ENT), angular second moment (ASM) and correlation (COR). "In some studies, homogeneity is called inverse different moment, contrast is called inertia and, angular second moment is called energy or uniformity" (Shi, 2003). The equations for the eight texture features applied in this study are in Equation 4.

$$HOM = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}, \quad SD = \sigma_I = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}\,(i-\mu_I)^2},$$

$$CON = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}\,(i-j)^2, \quad ENT = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} -P_{i,j}\,\ln P_{i,j},$$

$$DIS = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}\,|i-j|, \quad ASM = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}^2,$$

$$MEAN = \mu_I = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i\,P_{i,j}, \quad COR = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{(i-\mu_I)\,(j-\mu_J)}{\sigma_I\;\sigma_J},$$

Where N is the number of grey levels; P is the normalized symmetric GLCM of dimension N x N and Pi,j is the (I,j)th element of P;

$$P; \quad \mu_J = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} j\,P_{i,j}; \quad \text{and} \quad \sigma_J = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{i,j}\,(j-\mu_J)^2}$$

Equation 4. Equations of the eight GLCM texture features (After Shi, 2003)

"HOM measures local homogeneity, and results in a large value if the elements of the GLCM are concentrated on the main diagonal. CON measures local spatial frequency; if the GLCM has large off-diagonal elements, the local window has high contrast. DIS is similar to CON—high contrast of the local window indicates high DIS value. MEAN and SD measure the mean and standard deviation in terms of the GLCM. ENT measures disorder of the image, while ASM indicates local uniformity" (Shi, 2003). This explanation by Shi helps the user understand the intended function of each texture feature. The main part of the study was to see what effect adding more and more texture features together. Figure 13 describes the methodology used.

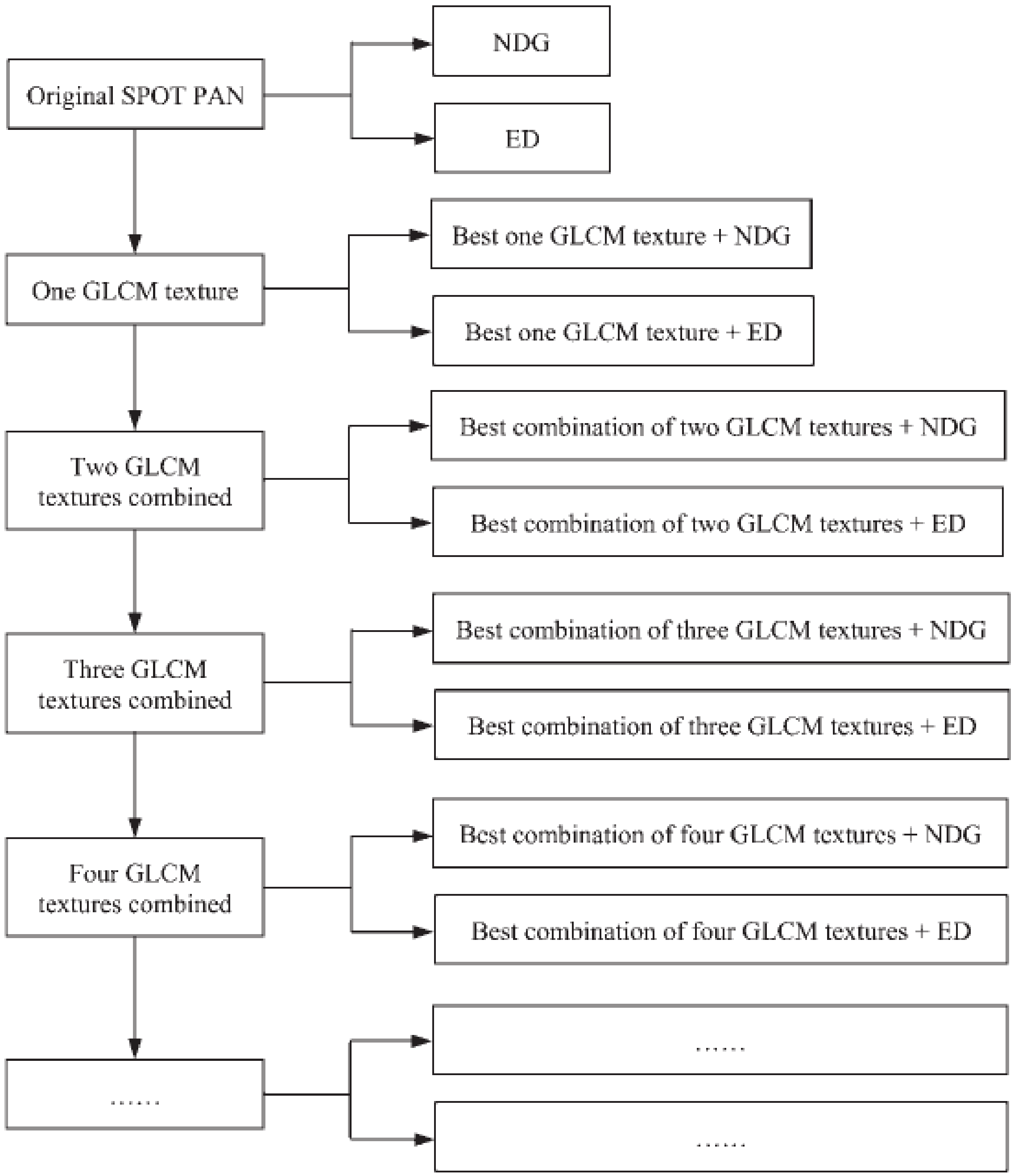


Figure 13. Texture Feature Methodology (From Shi, 2003)

While not all results are shown in this thesis, the same methodology of starting with just the pan image and adding texture features is used. The overall accuracies of just the texture feature of the SPOT images are shown in Table 1.

Input channels	Overall accuracy (%)	Kappa × 100
Original SPOT PAN	27.0	16.6
One GLCM texture feature		
DIS	37.8	27.5
HOM	37.3	26.7
SD	37.1	25.4
ASM	35.2	22.6
ENT	30.5	19.9
CON	30.2	20.0
MEAN	28.9	18.9
COR	16.7	5.9
Combinations of two GLCM texture features (top five combinations)		
MEAN+ENT	59.3	52.7
MEAN+DIS	58.8	51.8
MEAN+HOM	58.5	51.1
MEAN+SD	56.6	49.6
MEAN+CON	56.0	48.8
Combinations of three GLCM texture features (top five combinations)		
MEAN+ENT+COR	67.3	61.8
MEAN+DIS+COR	67.3	61.8
MEAN+DIS+SD	66.8	61.0
MEAN+HOM+COR	66.6	60.9
MEAN+HOM+SD	66.3	60.5
Combinations of four GLCM texture features (top five combinations)		
MEAN+HOM+SD+COR	70.2	65.0
MEAN+SD+ENT+COR	69.6	64.2
MEAN+CON+ENT+COR	69.3	64.0
MEAN+HOM+DIS+COR	69.2	63.9
MEAN+DIS+SD+COR	69.0	63.5
Combinations of five GLCM texture features (top five combinations)		
MEAN+SD+ENT+ASM+COR	71.3	65.0
MEAN+HOM+SD+ASM+COR	71.3	64.2
MEAN+DIS+SD+ASM+COR	70.9	64.0
MEAN+CON+ENT+ASM+COR	70.8	63.9
MEAN+CON+DIS+ASM+COR	70.7	63.5

Table 1. Results of just texture features on un-stratified SPOT image (From Shi, 2003)

The results show that the more texture features added, the more overall accuracy increases. The overall accuracy made bigger gains in the first couple of texture features and leveled off when five were added together. This aspect will be looked at for this thesis. Within the Shi study, the Hall–Beyer 2000 study was referenced, which divided the eight GLCM texture features into three groups: the contrast group (CON, DIS and HOM), the orderliness group (ASM and ENT), and the descriptive statistics group

(MEAN, SD and COR). The texture features in the contrast group are correlated with each other; so are the features in the orderliness group. MEAN and COR are generally not correlated with other features. Hall-Beyer suggested that using a texture feature from each group would help maximize results for classification purposes. Also, another point to be noted is that texture features performed differently in texturally different regions in the study area. For more homogeneous regions, single or combinations of two texture features had better performance, and fewer numbers of texture features were needed to approach the peak of classification accuracy.

4. Radar Altimeter Mean Return Waveforms from Near-Normal-Incidence Ocean Surface Scattering (Hayne, 1980)

A 1980 study by Hayne discusses the radar response for reflections off the surface of the ocean. The study describes skewness that is an occurrence texture feature. The results of this study offer promise in defining the impact of sunglint in optical data. The elimination of sunglint is the biggest factor in being able to detect just the kelp; skewness could help in the understanding of the phenomena of sunglint. Most early radar studies just assumed a simple Gaussian probability distribution to describe the ocean surface. This assumption is made to simplify the calculations. Hayne’s study describes the ocean more accurately. The study includes skewness and kurtosis, which are the normal distributions third and fourth moments. Equation 5 shows the equation including skewness and kurtosis.

$$q_s(t) = \left[1 + \frac{\lambda_s}{6} H_3(t/\sigma_s) + \frac{\kappa_s}{24} H_4(t/\sigma_s) + \frac{\lambda_s^2}{72} H_6(t/\sigma_s) \right]$$

$$\times \frac{1}{\sqrt{2\pi}\,\sigma_s} \exp\left[-\tfrac{1}{2}\,(t/\sigma_s)^2\right], \qquad ($$

$$H_3(z) = z^3 - 3z$$

$$H_4(z) = z^4 - 6z^2 + 3$$

and

$$H_6(z) = z^6 - 15z^4 + 45z^2 - 15.$$

Equation 5. Gaussian probability distribution with Skewness and Kurtosis (From Hayne, 1980)

Equation 5 helps describe the mean return waveform with respect to time. This is based on the use of radar using a Gaussian antenna on the satellite. This thesis does not use radar, but this basically would be the case with the sun being the signal. The charts below describe the mean return waveform of ideal Gaussian radar. Figure 14 describes the difference in the waveform response with the changing of ocean wave height. The higher the wave height, the more sloped the response becomes. Figure 15 shows the effects of skewness. Figure 16 shows the effect of introducing kurtosis along with skewness.

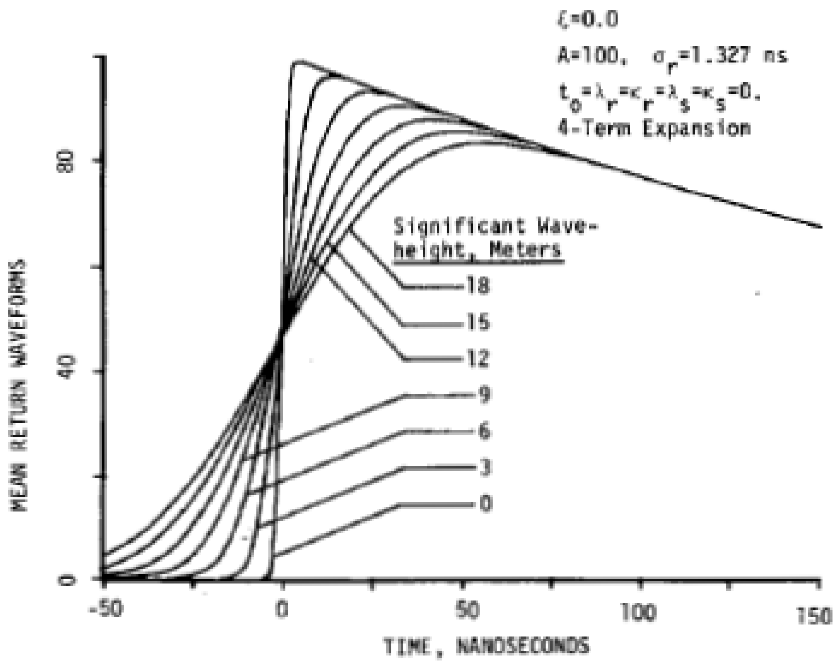


Figure 14. Idealized SEASAT radar altimeter mean return waveforms, showing effects of different ocean significant wave heights. (From Hayne, 1980)

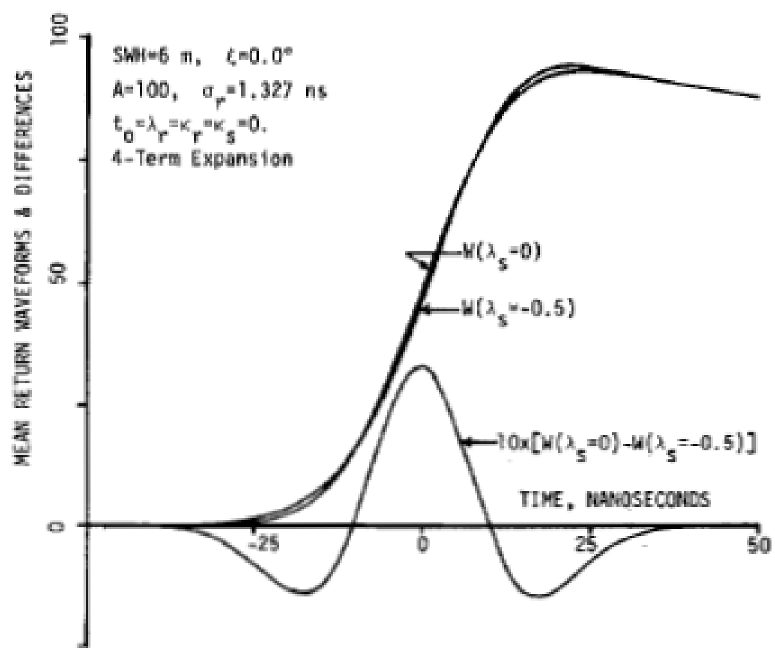


Figure 15. Idealized SEASAT radar altimeter mean return waveforms, showing effects of skewness in surface elevation probability density function. (From Hayne, 1980)

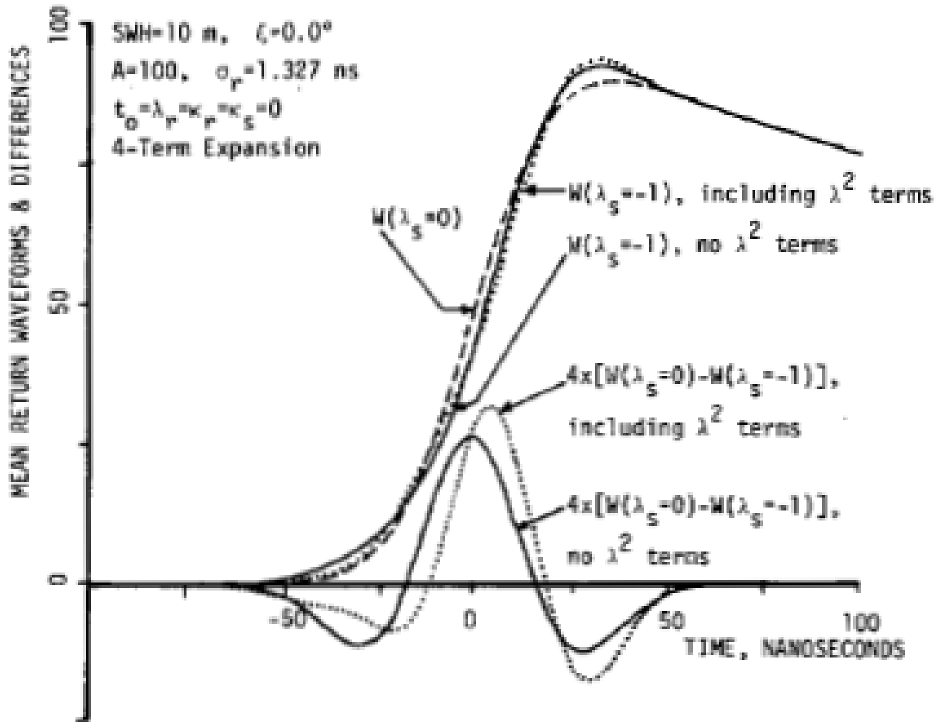


Figure 16. Idealized SEASAT radar altimeter mean return waveforms, showing effects of including skewness squared terms in surface elevation probability density function. (From Hayne, 1980)

E. QUICKBIRD SATELLITE

Figure 17. QuickBird Satellite (Picture taken from DigitalGlobe Web page, http://www.digitalglobe.com)

This project requires combined panchromatic and multispectral satellite data, such as those available from IKONOS and QuickBird. Quickbird, launched in 2001, provides sub-meter panchromatic imagery, and 2.4 meter multispectral data. QuickBird collects panchromatic imagery at 60–70 centimeter resolution and multispectral imagery at 2.4–2.8 meter resolutions (QuickBird Products Imagery Products Guide, 2009). The imagery can be imported into remote sensing image processing software such as ENVI for analysis. The panchromatic and multispectral imagery are collected simultaneously.

F. ENVI SOFTWARE

The Environment for Visualizing Images (ENVI) software was used to process the images in this thesis. For this thesis, ENVI's basic image manipulation tools were used to prepare the images for processing. The 13 co-occurrence and occurrence texture filters were used along with the Maximum Likelihood, Binary Encoding, Mahalanobis Distance, and Minimum Distance classification functions. The confusion matrices and Receiver Operating Characteristics (ROC) curve functions were used to create the results. Spectral analysis calculations, such as the NDVI calculation, are done using the standard ENVI tools.

III. OBSERVATIONS

A. INTRODUCTION

The goal of this work is to extend our ability to detect kelp into the domain of panchromatic imagery, for those cases where multispectral data are not available. The approach described here requires modest adjustment in the registration of the panchromatic and spectral data.

B. DATA SET

The data used for this analysis are from the Quickbird satellite. These data were collected at 18:38.22 UT, on September 5, 2003. Figure 18 illustrates the two sets.

Figure 18. MSI image on left and Panchromatic Image on right

C. INITIAL PROCESSING

The comparison of the panchromatic and multi-spectral images was completed using the following process.

First, the pan image (PAN: samples 6912, lines 7168) needed to be resized by a factor of 4 to match the size of the multi-spectral image (MSI: samples 7168, lines 7168).

Basically, the MSI received a black strip on the right side of the image to make square but no added value information. Figure 19 is a picture of the Santa Barbara coastline.

Then, the pan image needed to be warped to match the MSI image. The warping of the pan image was done by taking 10 to 12 Ground Control Points (GCPs) by matching pixels from each image and linking them together, and then warping the image.

The land is then masked because we are only concerned with vegetation in the water. To do this, Region of Interest (ROI) is created over the land. This Land ROI is used to set all the values for land to zero. Now, only objects in the ocean that reflect the designated wavelengths will have positive values. These reflections could be caused by anything in the ocean such as sunglint, kelp, and ships. Since the Pan and MSI images are now of the same spatial size, the Land ROI is applied to both images.

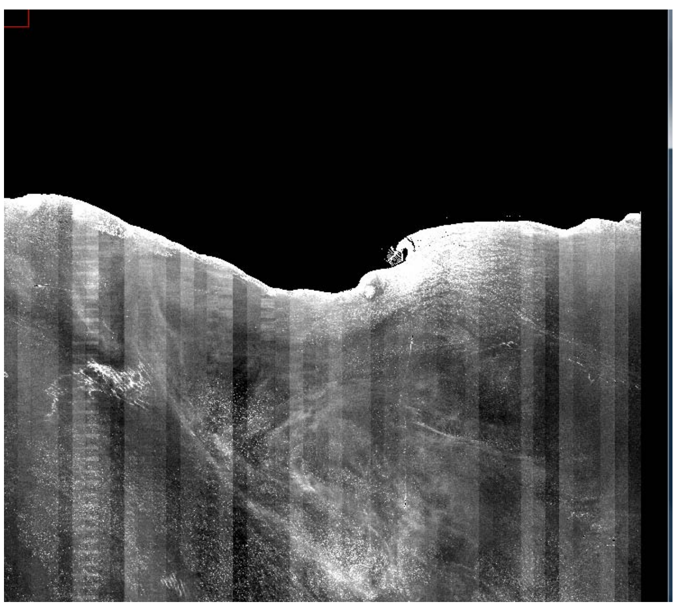


Figure 19. Masked Pan image

Next, the Normalized Difference Vegetation Index (NDVI) will be calculated using the Multi-Spectral Image. This process is made easier by a function contained in the ENVI 4.5 software. There is a pre-loaded function that converts the values of:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

Equation 6. Normalized Difference Vegetation Index (NDVI)

Figure 20. NDVI

The NDVI and Pan images are now ready to be compared. The NDVI image, shown in Figure 20, is the truth or reference image. The Pan image will be analyzed by the intensity comparison, texture features, confusion matrices, with analysis of the results by use of ROC curves sections.

The first element of the analysis that follows will be an exploration of small sections of the scene to determine how good the relationship is between NDVI and simple brightness in the panchromatic data.

Secondly, the eight co-occurrence and five occurrence texture features are calculated for the Pan image and compared to the NDVI image by use of two-dimensional scatter plots. This comparison is conducted to see how well the texture feature distinguished the kelp from water. A 3×3 search window was used and the size of a search window usually corresponds to the size of the object that is being evaluated. In this case, it will be the kelp. Since the original PAN image needed to be warped, the warped Pan image has a reduction in resolution by 4 times. This was done in order to compare it to the truth NDVI image. If the texture analysis was done to the original image, a search window of 12×12 would be to be use to produce comparable results.

Thirdly, confusion matrices will be done on the pan, variance, and classification matrices. This will also include the SBC-LTER Research Area of Arroyo Burro and Mohawk. For the pan and variance images, a threshold value will be used to create each region of interest. The higher values will represent the kelp and the lower values will represent the water. The classification methods with the corresponding texture feature will also create kelp and water regions of interest. These results, along with the truth image, will be used to create the confusion matrices.

Lastly, the pan and variance images will be used to create ROC curves. These curves show how well a variable correctly classifies an object.

IV. OBSERVATIONS AND ANALYSIS

A. OBSERVATION AND ANALYSIS FORMAT

For the observation and analysis section, there are four different sections comprised of the intensity comparison, texture features, confusion matrices, and ROC curves as described in the process section.

B. INTENSITY COMPARISON

It is time to see how accurately the NDVI and PAN images compare.

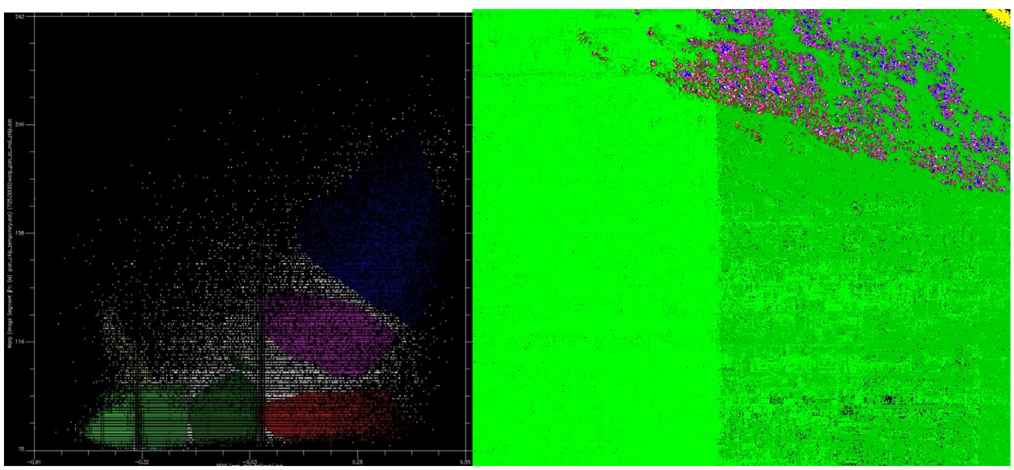


Figure 21. Regions of kelp and water

In Figure 21, the left side of the figure is a two-dimensional scatter chart with NDVI on the x axis and PAN on the y axis. Each point on the chart are the values of the pixels in NDVI and PAN. The location in the chart of each point is (NDVI, PAN). The values of the NDVI axis range from -1 to 1, while the PAN axis ranges from 0 to 255. This chart is used to see how the PAN image correlates with the NDVI image. Linear patterns of highly correlated points could show that there is a relationship between the NDVI and Pan image. If there is a relationship, this could be used to distinguish the kelp from the water in the Pan image without the help of the NDVI image. An area is selected such as the one selected on right side of Figure 21, then the 2D Scatter Chart function in the ENVI 4.5 software is used to produce the scatter chart.

In the scatter chart, a region of points can be highlighted with any color. These different colors sections of the scatter chart represent the kelp, water and sunglint. While all of the pixels will be classified as kelp or water, sunglint is a subset of of water, and it is important to characterize this phemona to be able to discriminate it from the kelp. How well the NDVI and Pan image correlate in this scatter plot will give the starting point to see whether Pan imagery can be used to classify kelp.

Kelp lies between 0 and 1, and water lies between -1 and 0 on the NDVI axis. The blue and purple regions show the linear correlation of the kelp between the NDVI and PAN images. This is shown by the diagonal line on the right side of the scatter plot. The green area represents the area correlated as water in both the NDVI and PAN images. This is where the bulk of the pixels are located, even though it looks more spread out in the kelp portion of the scatter plot. This is a very important characteristic of the plot, as shown in the right side of Figure 21, where almost the whole image is green. Kelp beds are relatively small compared to the amount of area that is water. This means there is much more area that could have sunglint in it compared to the kelp. The image on the right of Figure 21 shows how much area is covered by each color region. The area in red in the most interesting part of the scatter plot in regards to the correct classification of the kelp and water. It is the area of low PAN intensity that has a high NDVI rating. This area would not be able to be correctly detected in a simple threshold of the PAN image. This means that if the Pan image was cut if half by a predetermined number, the higher values would be kelp and the lower values would be water. In those lower values of water, there would be kelp incorrectly classified. Overall, the scatter plot does show that a simple threshold is a pretty accurate classification of the kelp. With more analysis, exactly how well it does can be defined.

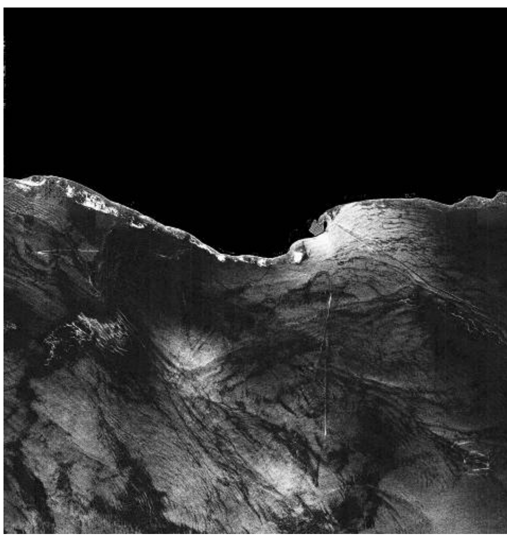
C. TEXTURE FEATURES

To better understand the texture features that will used to classify the kelp and water in the classification matrices, the images of each texture feature is shown along with the two dimensional scatter charts. The texture feature image is just the resulting values that can vary in ranges depending on each equation. The 2D Scatter chart is the same form as in Figure 21, with NDVI on the x axis and the texture feature on the y axis.

1. Occurrence

The occurrence texture features are used here to create an image of each texture feature and a two-dimensional scatter chart. The occurrence texture features are first-order texture features.

a. Data Range



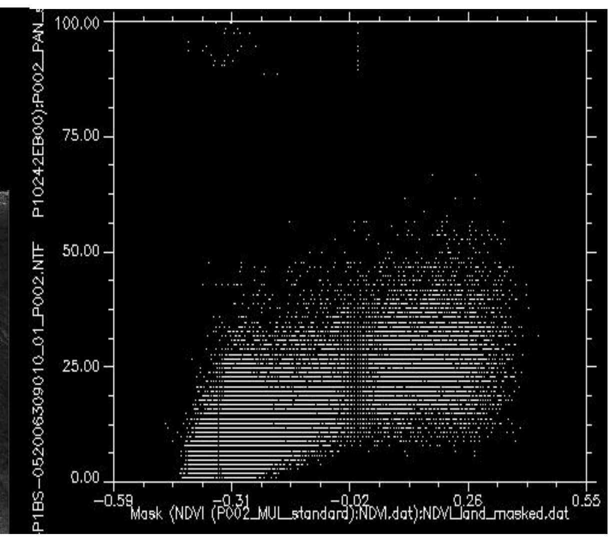


Figure 22. Data Range Image and (NDVI, Data Range) 2D scatter plot

The data range texture feature is simple to describe because it is just the difference in values of the pixels. The gray tone values that can range from 0 to 255, but the pan image ranges from 75 to 150. Knowing that most of the water is relatively constant, the vast majority of the points on the 2D scatter chart in Figure 24 are located between -0.4 and -0.1 on the x axis and near zero on the y axis. The rest of the data range

2D scatter plot is the interesting part of the chart and does not show or have any specific statistical information. Since there is about the same value for the difference between water and sunglint and the kelp and water, the data range texture feature doesn't have the ability to distinguish the kelp and sunglint.

b. Mean

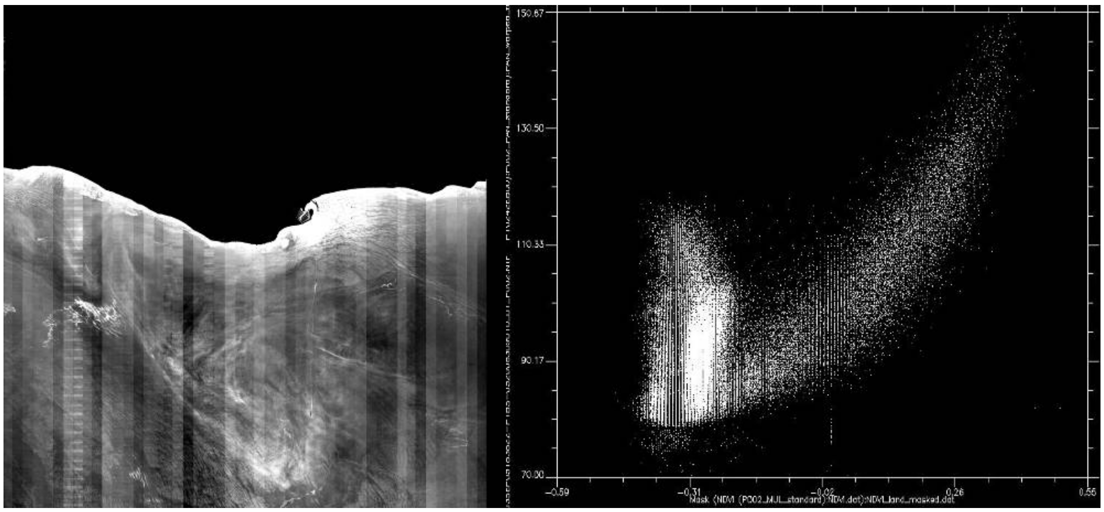


Figure 23. Occurrence Mean Image and (NDVI, MEAN) 2D scatter plot

The occurrence mean texture feature provides results similar to those of the co-occurrence mean texture feature. Since it is the mean, the outliers will be pulled in and tighten any pattern in the Pan image. Taking a closer look produces some differences between these mean texture features. The co-occurrence mean values range from 0 to 10, while the occurrence mean values range from 75 to 150 much like the pan image values. This makes sense because the occurrence mean image looks much like the pan image.

c. Variance

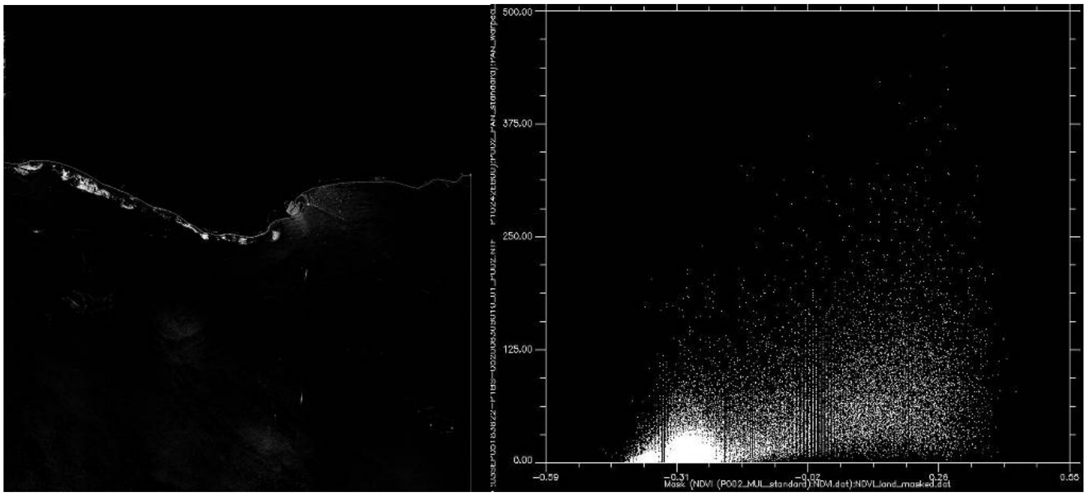


Figure 24. Occurrence Variance Image and (NDVI, Variance) 2D scatter plot

The occurrence variance texture feature provides results similar to those of the co-occurrence variance texture feature. This variance 2D scatter plot does have an interesting ball of points for the water located lower than 50 on the y axis and between -0.4 and -2 on the x axis. As stated before, the kelp appears in the form of kelp beds have mostly higher values but don’t seem to have a really high variance within the kelp bed, but it is noticeably higher than the water. This is a factor that could be used to distinguish the kelp from water. In the occurrence variance image on the left side of Figure 24, the area along the coastline have brighter spots which represent the kelp and the higher values on the 2D scatter plot.

d. Entropy

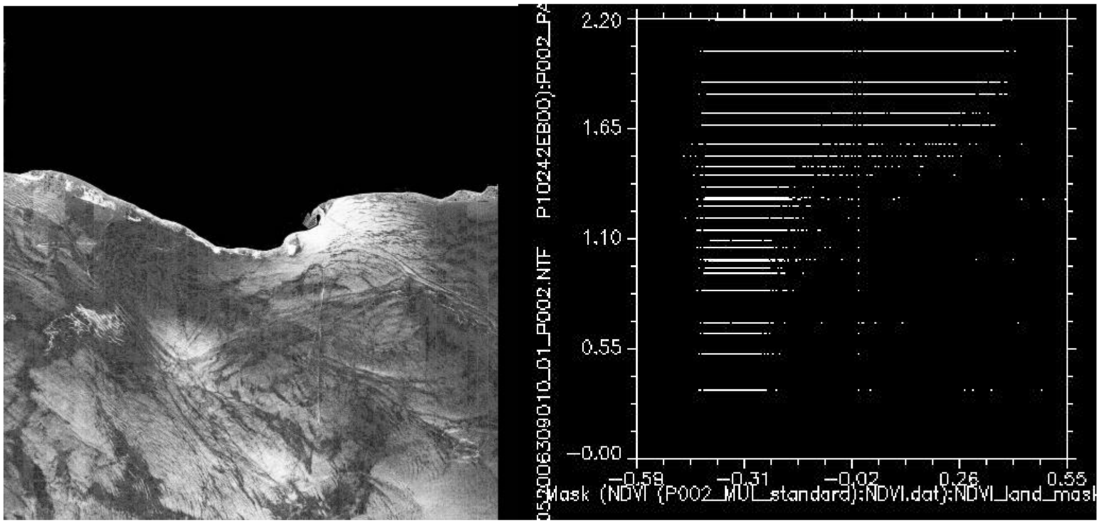


Figure 25. Occurrence Entropy Image and (NDVI, Entropy) 2D scatter plot

The occurrence entropy texture feature produces a very unique image and 2D scatter chart, as shown in Figure 25. Most of the of the other texture feature images have shown the water area to be mostly a solid color with a little distingishng areas where the kelp is located near the coast. In this case, there are swirls all over the area of the water. How does this translate to the occurrence entropy 2D scatter chart? The water and sunglint range throughout the entire y axis and between -0.4 and 0 on the x axis. The kelp is only in the upper regoins of the y axis between 2.2 and 1.5. Since the water shares the higher values along with the kelp, this will not be useful in discrimnating the kelp from the water.

e. Skewness

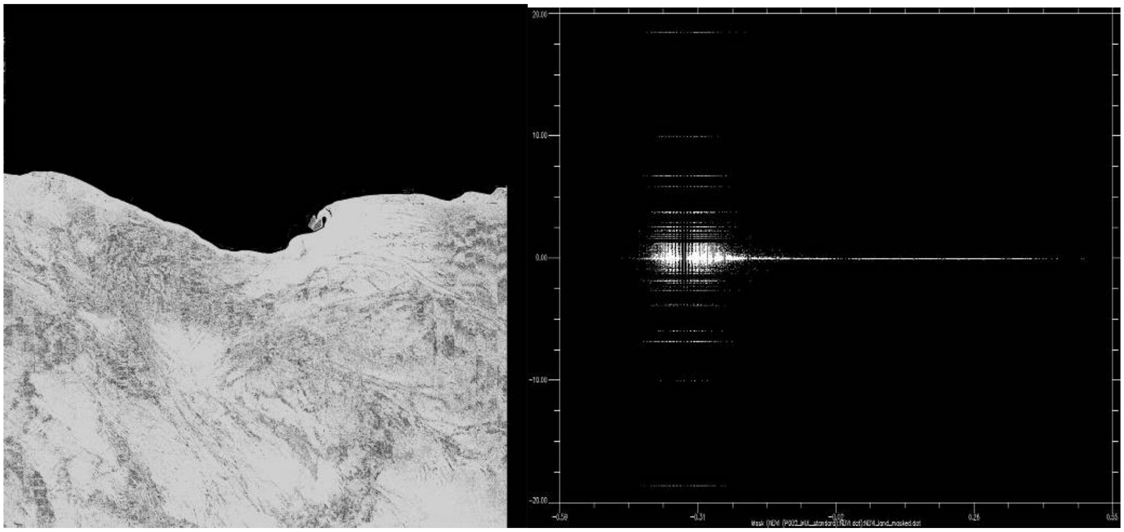


Figure 26. Skewness Image and (NDVI, Skewness) 2D scatter plot

The skewness texture feature measure the degree at which a normal distribution deviates to the left or right. This is much like a wave in the ocean. As the wave starts to curl over, the skewness starts to increase in value. A normal distribution would produce a result of zero. In the skewness 2D scatter plot on the right in Figure 26, the range of the kelp the skewness levels are near zero. In the range of the water, there are lines of positve and negative skewness. When looking at the skewness image on the left side of Figure 26, the waves are very distictive and the regions of the kelp are very smooth. After selecting the value of zero and near zero, the low intesity kelp cannot be extracted. The thought was that skewness could be use to pull out the sunglint from the image, but that does not seem to be the case.

2. Co-occurrence

The second-order statistics are calculated using a 3×3 window, for steps of 1 pixel in the X and Y directions.

a. Mean

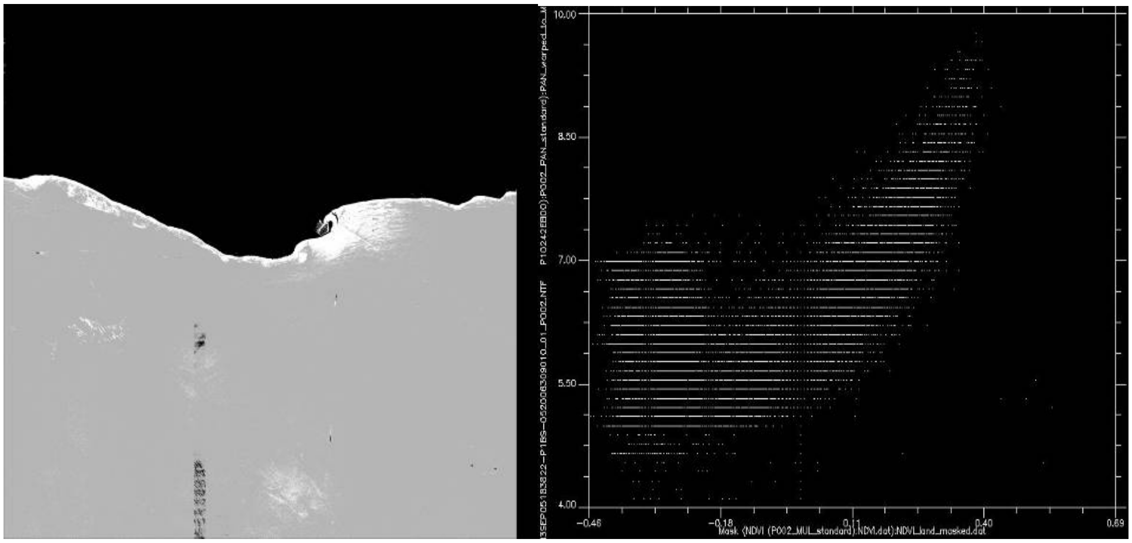


Figure 27. Mean Image and (NDVI, MEAN) 2D scatter plot

The mean texture feature averages the area for each Gray-Tone Spatial-Dependence Matrices. Since it is the mean, the outliers will be pulled in and tighten any pattern in the Pan image. The mean 2D scatter chart is similar to the Pan 2D scatter chart in Figure 27 with the additional tightening of the values. The scatter chart on the right shows area cluster on points to the left, which is the water pixels. Then, on the right side between 0 and 1 on the x axis, the mean has a linear relationship of the kelp. The values extend much higher on the kelp, which shows a distinctive characteristic of the kelp. This relationship should prove useful in the classification of the kelp and water.

b. Variance

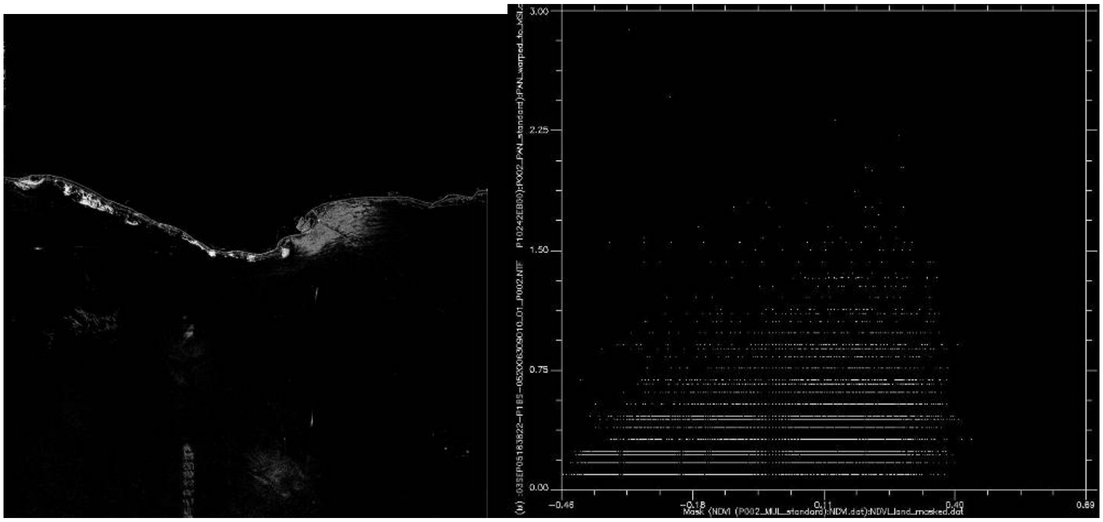


Figure 28. Variance Image and (NDVI, Variance) 2D scatter plot

The variance texture feature will give higher values for the areas with large differences between the pixels next to each other in each Gray-Tone Spatial-Dependence Matrices. Water is a terrain that does not have much variance except for the sunglint. The kelp beds have mostly higher values but also not much variance within the kelp bed; however, it is noticeably higher than the water. This is a factor that could be used to distinguish the kelp from water. There is something present that is hard to see in the scatter plot that can be shown by the image on the left. A large number of pixels have a very low variance represented by the large amount of black in the variance image. On the 2D scatter chart, the points are so close to the bottom that they cannot be seen. The areas where the kelp is located have the lighter spots, which represent higher values.

c. Homogeneity

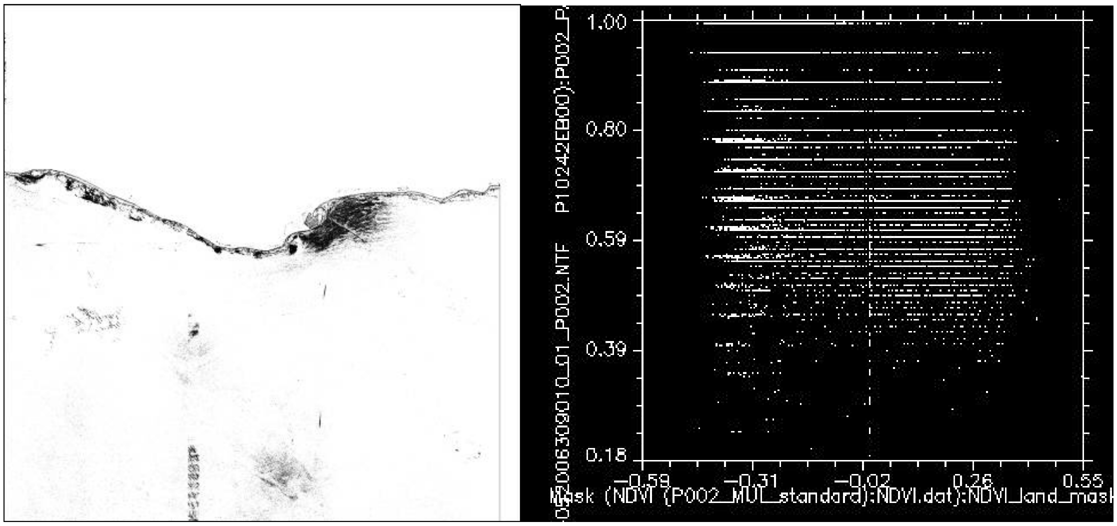


Figure 29. Homogeneity Image and (NDVI, Homogeneity) 2D scatter plot

The homogeneity texture feature gives higher values for areas that are more uniform. The 2D scatter chart shows a thin line between -0.4 and 0 on the x axis and at the values of 1 on the y axis. This is the large number of water pixels that can also be shown in the in all of the white area in the homogeneity image on the left hand side of Figure 29. In the 2D scatter plot, areas of lower homogeneity the sunglint and kelp are randomly scattered from -0.4 to 0.4 on the x axis. Without any discrimination between these, this does not give the user any specific statistical information that would help classify the water and kelp.

d. Contrast

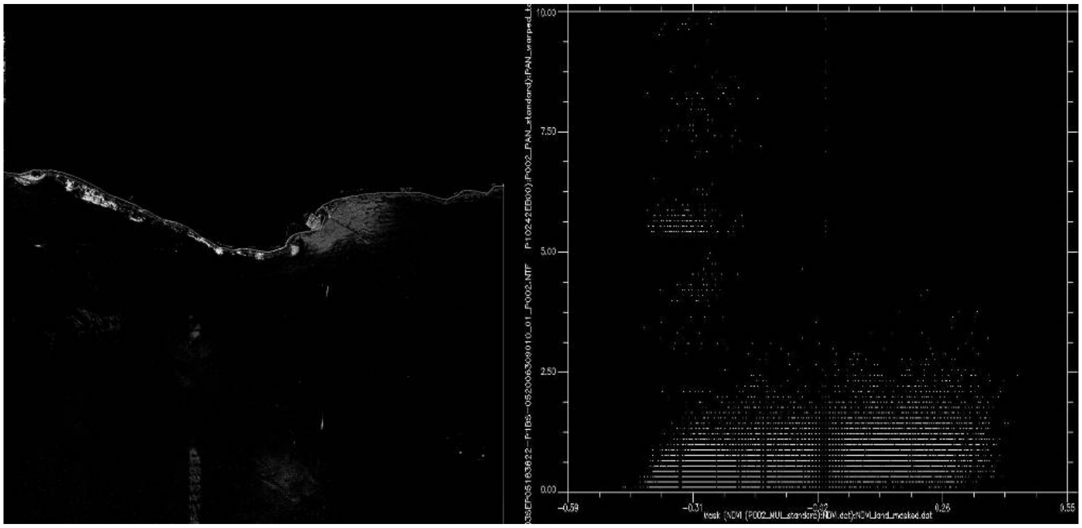


Figure 30. Contrast Image and (NDVI, Contrast) 2D scatter plot

The contrast texture feature will give higher values for the areas with larger differences between pixels within each Gray-Tone Spatial-Dependence Matrices. This is similar to the variance but is calculated a little differently. The contrast 2D scatter plot and image as would be expected is similar to variance. Until contrast and variance are used in the classification methods, the extent to which these are similar cannot be further described here.

e. Dissimilarity

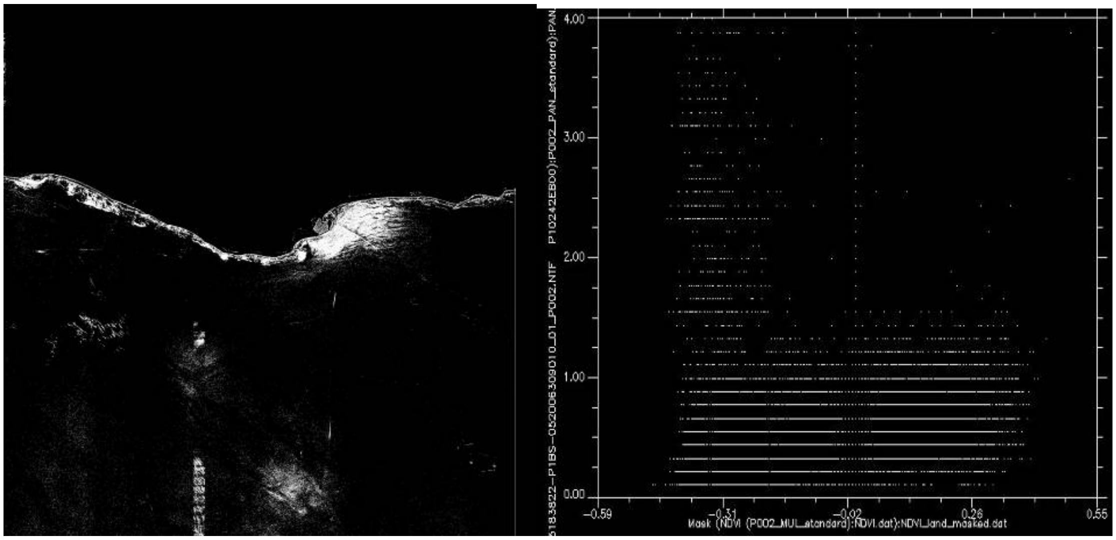


Figure 31. Dissimilarity Image and (NDVI, Dissimilarity) 2D scatter plot

The dissimilarity texture feature will give higher values for the areas with larger differences between pixels within each Gray-Tone Spatial-Dependence Matrices. This is similar to and is calculated very closely the contrast. For dissimilarity, the sunglint and kelp cannot be distingished from each other, so it is not likely that it would not help classify the water and kelp correctly.

f. Entropy

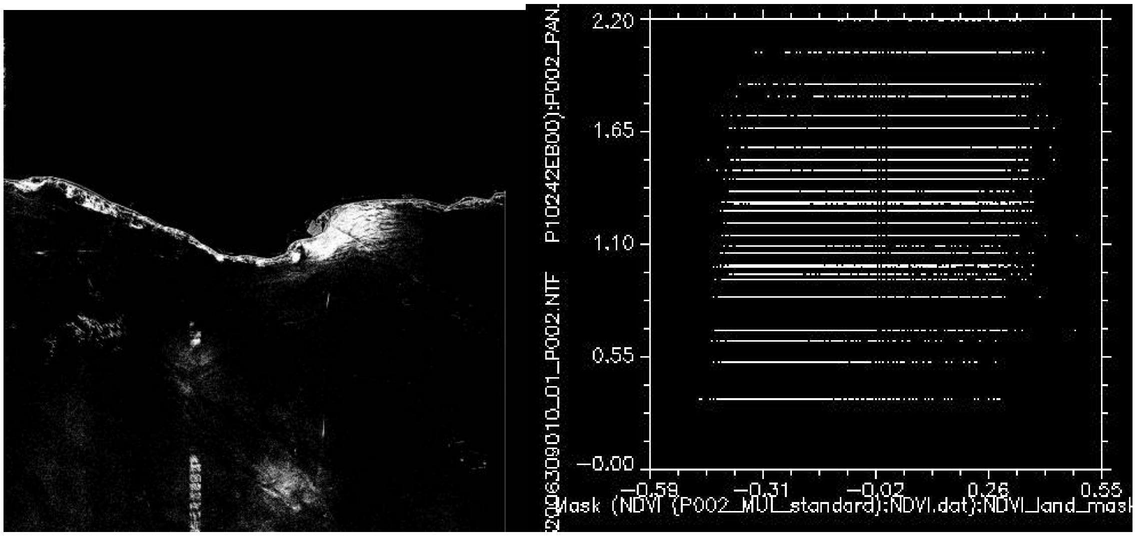


Figure 32. Entropy Image and (NDVI, Entropy) 2D scatter plot

The entropy texture feature will give higher values for the areas with larger differences between pixels within each Gray-Tone Spatial-Dependence Matrices. A way to think about entropy is the more chaos, the higher the value for the entropy. Entropy is similar to contrast and dissimilarity where a bulk of the water is valued at zero and represented by the large amount of black in the image of Figure 32. This entropy 2D scatter plot shows random values from -0.4 to 0.4. The sunglint and kelp cannot be distingished from each other, so this would not help discriminate between the water and kelp.

g. *Second Moment*

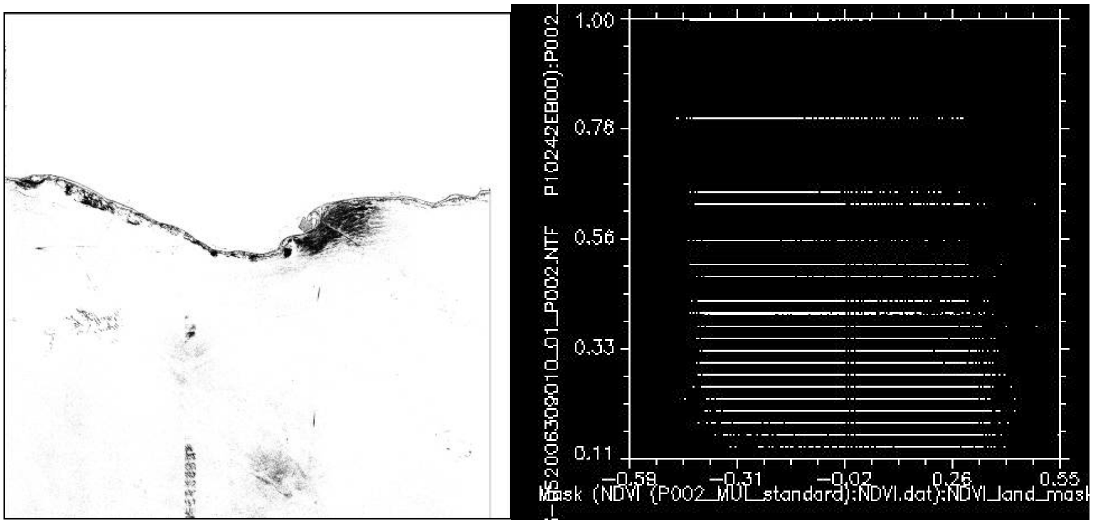


Figure 33. Second Moment Image and (NDVI, Second Moment) 2D scatter plot

The second moment texture feature gives higher values for areas that more uniform. The 2D scatter chart shows a thin line between -0.4 and 0 on the x axis and at the values of 1 on the y axis just as in the homogeneity chart. This is the large number of water pixels that can also be shown in all of the white area in the second moment image on the left side of Figure 33. Just like the homogeneity 2D scatter plot, areas of lower second moment values, which are the sunglint and kelp, are randomly scattered from -0.4 to 0.4 on the x axis. Without any discrimination between these, this does not give the user any specific statistical information that would help classify the water and kelp.

h. Correlation

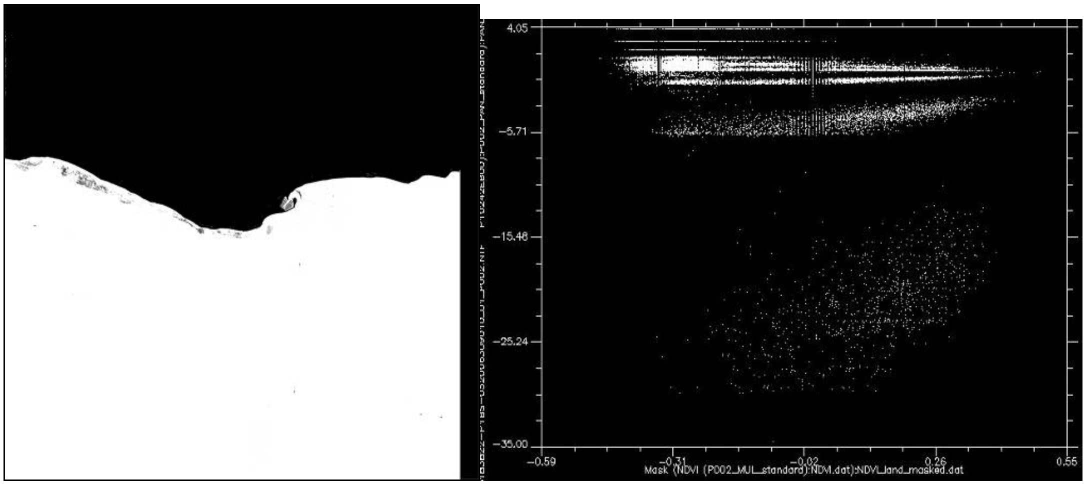


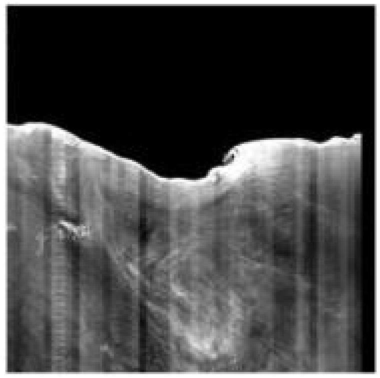
Figure 34. Correlation Image and (NDVI, Correlation) 2D scatter plot

This correlation 2D scatter plot gives higher values to the areas with linear-dependencies. Since the different bands range across the NDVI values in the 2D scatter plot, the bands would not help in distinguishing between the kelp and water.

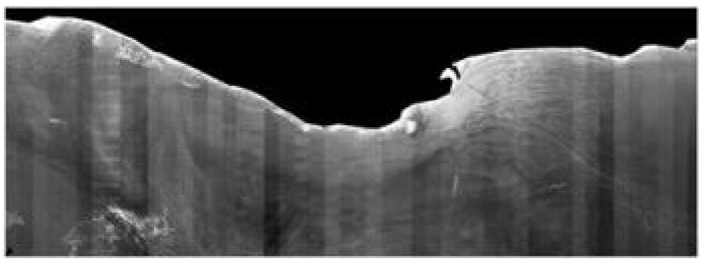
D. CONFUSION MATRICES

Confusion matrices show how many pixels are correctly and incorrectly classified. This information is produced by comparing a set of objects to a truth image. The truth image in this thesis is the NDVI image. There can be as many objects to classify as the user wants. This thesis only has two objects to classify, the kelp and water. There needs to be some method to classify the set of objects. There are two kinds used, which are simple thresholds and classification methods.

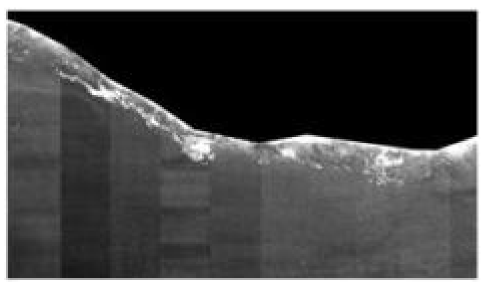
There are three images seen in Figure 35: large, small, and study area used for the analysis. The large image has a vast search area compared to the amount of kelp. This area has a high potential for sunglint compared to kelp. The small image reduces the water coverage while maintaining the coast line where the kelp is located. The study area image will provide coverage and results for the SBC-LTER area.



(a) Large Image



(b) Small Image



(c.) Study Area Image

Figure 35. Images used for Analysis (a, b, and c)

In the following confusion matrices, the PAN and Variance matrices use a simple threshold method for classification of the kelp and water. There are two very important pieces of information that are being sought in these confusion matrices. The first is the quantitative analysis of the sunglint. Most of the water has a very low value and can be correctly classified by low threshold value. When the threshold is done, the number of sunglint pixels can be determined. With most all of the gray tone values for the Pan image between 75 and 150, steps of five in gray tone values were chosen to determine how the sunglint and kelp responded to each increase in value. The variance texture feature values are dependent on the Pan image, since they are calculated based on it. The values and steps in values were chosen after some analysis of the 2D scatter chart and the range in values. The second piece of information is the simple thresholds set a baseline to evaluate the classification methods with the selected texture features. Any classification method that does not at least meet the simple threshold adds no value.

There are four classification methods that are used which are Maximum Likelihood, Mahalanobis Distance, and Minimum Distance, and Binary Encoding. The Maximum Likelihood classification sets parameters for each classification type such as kelp and water in this case. All pixels are given a probability of being in each class and the class with the highest probability is what the pixel will be assigned (Richards, 1999). The Mahalanobis distance classification is a direction-sensitive distance classifier that is similar to the maximum likelihood classification but assumes all class covariances are equal. All pixels are classified to the closest classification type (Richards, 1999). The Minimum Distance classification method uses the mean vectors of each endmember, which are pure spectrally unique materials of each classification type like the kelp and water. Then, it calculates the Euclidean distance from each unknown pixel to the mean vector for each class. Then, the pixels are classified to the nearest classification type (Richards, 1999). While these three methods are related, the Binary Encoding is a bit different. The Binary Encoding classification method encodes the data and endmember spectra into zeros and ones, based on whether a band falls below or above the spectrum mean, respectively. An exclusive OR function compares each encoded reference spectrum with the encoded data spectra and produces a classification image. All pixels are classified to the endmember with the greatest number of bands that match (Mazer, 1988).

1. Large Pan Image

		Ground Truth (Pixels)			Ground Truth (%)		
Gary Tone Value	feature	kelp NDVI	Water NDVI	total	kelp NDVI	Water NDVI	total
80	kelp	59126	13869722	13928848	99.97	48.53	48.63
OA = 51.5745%	water	19	14712541	14712560	0.03	51.47	51.37
KC = .0044	total	59145	28582263	28641408	100	100	100
85	kelp	59108	1795885	1854993	99.94	6.28	6.48
OA = 93.7296%	water	37	26786378	26786415	0.06	93.72	93.52
KC = .0580	total	59145	28582263	28641408	100	100	100
90	kelp	58588	635879	694467	99.06	2.22	2.42
OA = 97.7779%	water	557	27946384	27946941	0.94	97.78	97.58
KC = 0.1523	total	59145	28582263	28641408	100	100	100
95	kelp	56789	292507	349296	96.02	1.02	1.22
OA = 98.9705%	water	2356	28289756	28292112	3.98	98.98	98.78
KC = 0.2755	total	59145	28582263	28641408	100	100	100
100	kelp	53090	142121	195211	89.76	0.5	0.68
OA = 99.4827%	water	6055	28440142	28446197	10.24	99.5	99.32
KC = 0.4156	total	59145	28582263	28641408	100	100	100
105	kelp	47660	74052	121712	80.58	0.26	0.42
OA = 99.7014%	water	11485	28508211	28519696	19.42	99.74	99.58
KC = .5257	total	59145	28582263	28641408	100	100	100
110	kelp	41515	36937	78452	70.19	0.13	0.27
OA = 99.8095%	water	17630	28545326	28562956	29.81	99.87	99.73
KC = 0.6025	total	59145	28582263	28641408	100	100	100
115	kelp	35349	14861	50210	59.77	0.05	0.18
OA = 99.8650%	water	23796	28567402	28591198	40.23	99.95	99.82
KC = 0.6458	total	59145	28582263	28641408	100	100	100
120	kelp	29716	4089	33805	50.24	0.01	0.12
OA = 99.8830%	water	29429	28578174	28607603	49.76	99.99	99.88
KC = 0.6389	total	59145	28582263	28641408	100	100	100
125	kelp	24484	1540	26024	41.4	0.01	0.09
OA = 99.8736%	water	34661	28580723	28615384	58.6	99.99	99.91
KC = 0.5744	total	59145	28582263	28641408	100	100	100

Table 2. Large PAN Image Confusion Matrices from threshold classification

With no gray tone values below 75, a value of 80 was selected for the first confusion matrix. The overall accuracy was 51.75% for 80, but with most of the water in the low 80s, the overall accuracy jumped to 93.73%. At this time, there is a very important point to bring up with the confusion matrices and the relative size of the water to the kelp. The numbers of truth NDVI kelp pixels are 59145 compared to the water NDVI pixels, which are 28582263. The water is the dominate factor in the overall accuracy of the image. In the 90 PAN confusion matrix, the PAN kelp has 694467 pixels, while there are only 59145 truth kelp pixels. This is only an 8.52% User Accuracy if one is just trying to find the kelp. But since the area is classified by both kelp and water, the water has a 97.78% User Accuracy, which brings up the overall accuracy.

These steps in the intensity really give the user a good characterization of the kelp and the sunglint. The kelp is 99% at 90 and 96% at 95, and then starts a trend of 10% drop as each 5 gray tone value goes up. This shows the linear relationship between the PAN intensity values and number of kelp. Also, the matrices characterize the water that has a bright reflection known as sunglint. From 90, where the sunglint is 635K, it reduces by half each as each 5 gray tone value goes up. When the sunglint is at 120, it is down to a meager 4K. After that, much of the only bright objects left are the kelp. This information will help in trying to assign additional attributes to the kelp and sunglint, which will help classify both of them more accurately.

2. Small Pan Image

Gary Tone Value	feature	Ground Truth (Pixels) kelp NDVI	Ground Truth (Pixels) Water NDVI	total	Ground Truth (%) kelp NDVI	Ground Truth (%) Water NDVI	total
80	kelp	58068	7952006	8010074	99.97	61.17	61.34
OA = 39.1058%	water	15	5048653	5048668	0.03	38.83	38.66
KC = .0056	total	58083	13000659	13058742	100	100	100
85	kelp	57326	1587780	1645106	98.7	12.21	12.6
OA = 87.8355%	water	757	11412879	11413636	1.3	87.79	87.4
KC = .0592	total	58083	13000659	13058742	100	100	100
90	kelp	54795	562425	617220	94.34	4.33	4.73
OA = 95.6679%	water	3288	12438234	12441522	5.66	95.67	95.27
KC = 0.1554	total	58083	13000659	13058742	100	100	100
95	kelp	51465	279279	330744	88.61	2.15	2.53
OA = 97.8107%	water	6618	12721380	12727998	11.39	97.85	97.47
KC = 0.2591	total	58083	13000659	13058742	100	100	100
100	kelp	47605	147430	195035	81.96	1.13	1.49
OA = 98.7908%	water	10478	12853229	12863707	18.04	98.87	98.51
KC = 0.3718	total	58083	13000659	13058742	100	100	100
105	kelp	43493	78427	121920	74.88	0.06	0.93
OA = 99.2877%	water	14590	12922232	12936822	25.12	99.4	99.07
KC = .4801	total	58083	13000659	13058742	100	100	100
110	kelp	39163	41535	80698	67.43	0.32	0.62
OA = 99.5371%	water	18920	12959124	12978044	32.57	99.68	99.38
KC = 0.5621	total	58083	13000659	13058742	100	100	100
115	kelp	34819	22844	57663	59.95	0.18	0.44
OA = 99.6469%	water	23264	12977815	13001079	40.05	99.82	99.56
KC = 0.5999	total	58083	13000659	13058742	100	100	100
120	kelp	30606	13330	43936	52.69	0.1	0.34
OA = 99.6875%	water	27477	12987329	13014806	47.31	99.9	99.66
KC = 0.6389	total	58083	13000659	13058742	100	100	100
125	kelp	26611	8440	35051	45.82	0.06	0.27
OA = 99.6944%	water	31472	12992219	13023691	54.18	99.94	99.73
KC = 0.5700	total	58083	13000659	13058742	100	100	100

Table 3. Small PAN Image Confusion Matrices from threshold classification

These confusion matrices in Table 3 represent the small pan image, which was band threshold exactly like the large PAN image, but the bottom of the water that is just empty space was eliminated to help in the large discrepancy between the numbers of water pixels as compared to kelp pixels. Another purpose of the image reduction is to see the effect on the amount of the sunglint pixels. The level of the sunglint pixels at the 90 level in the large PAN image was 635879 and the small PAN image 562425. With half of the water reduced, this showed that most of the kelp pixels remained in the smaller PAN image. So, the location of the sunglint was narrowed down in the small PAN image. Originally, sunglint was seen as one random occurrence that was caused by the reflection on the water, since the water flickers even within the waves and causes it to be random. This is true for most of the water in the ocean for the images.

However, there is a particular area where sunglint takes on a different dynamic. Near the shore and the harbor, there is white foam created breaking waves and the boats' wakes. This white foam tends to be harder to differentiate from the kelp. The white foam is more uniform and reflects a higher gray tone level like the kelp. The area of the shore is fairly large when compared to the size of the kelp beds. This is an area that would be needed to be eliminated in order to get a more accurate reading of the kelp. As for the small PAN image confusion matrices, they followed very closely to the large PAN image.

3. Variance Image

Variance Value	feature	Ground Truth (Pixels) kelp NDVI	Water NDVI	total	Ground Truth (%) kelp NDVI	Water NDVI	total
15	kelp	57848	876968	934816	99.6	6.75	7.16
OA = 93.2826%	water	235	12123691	12123926	0.4	93.25	93.84
KC = 0.1091	total	58083	13000659	13058742	100	100	100
30	kelp	57041	503154	560195	98.21	3.87	4.29
OA = 96.1390%	water	1042	12497505	12498547	1.79	96.13	95.71
KC = 0.1779	total	58083	13000659	13058742	100	100	100
70	kelp	51888	221114	273002	89.33	1.7	2.09
OA = 98.2593%	water	6195	12779545	12785740	10.67	98.3	97.91
KC = 0.3084	total	58083	13000659	13058742	100	100	100
100	kelp	47097	146183	193280	81.09	1.12	1.48
OA = 98.7964%	water	10986	12854476	12865462	18.91	98.88	98.52
KC = .3704	total	58083	13000659	13058742	100	100	100
150	kelp	39347	85918	125265	67.74	0.66	0.96
OA = 99.1986%	water	18736	12914741	12933477	32.26	99.34	99.04
KC = 0.4257	total	58083	13000659	13058742	100	100	100
200	kelp	32392	56219	88611	55.77	0.43	0.68
OA = 99.3728%	water	25691	12944440	12970131	44.23	99.57	99.32
KC = 0.4386	total	58083	13000659	13058742	100	100	100
250	kelp	26845	39630	66475	46.22	0.03	0.51
OA = 99.4573%	water	31238	12961029	12992267	53.78	99.7	99.49
KC = 0.4283	total	58083	13000659	13058742	100	100	100

Table 4. Variance Confusion Matrices

These confusion matrices in Table 4 represent the levels of variance in the image. A band threshold was conducted just like the PAN images. With the kelp gray tone levels higher as compared to the water, it was reasoned that the kelp pixels would be in the higher levels of variance. The PAN images were band threshold by a linear fashion in steps of five. The variance image was stepped by a doubling of the previous number, but the end was not quite doubled. This pattern was found after the first couple band thresholds and the number of kelp in each band. There are a few less band thresholds for the variance image, but it accurately describes the kelp within the difference bands. With almost all of the kelp accounted for in the PAN image, the number of miscounted pixels was 562425, while the number for the variance image was 503154. So, the variance image is better at ignoring the noise while picking out the kelp in the early band thresholds, but the higher levels show that the PAN image does better at the higher levels. A point to be made is that the kelp and water numbers are very near each other, and the overall accuracy of the images is very high.

4. Classification Image

		Ground Truth (Pixels)			Ground Truth (%)		
Classification	feature	kelp NDVI	Water NDVI	total	kelp NDVI	Water NDVI	total
BinEn - Pan Var	kelp	44215	162780	206995	76.12	1.25	1.59
OA = 98.6473%	water	13868	12837879	12851747	23.88	98.75	98.41
KC = .3289	total	58083	13000659	13058742	100	100	100
MahDis - Pan Var	kelp	44027	96256	140283	75.8	0.74	1.07
OA = 99.1553%	water	14056	12904403	12918459	24.2	99.26	98.93
KC = .4404	total	58083	13000659	13058742	100	100	100
MaxLike - Co-Oc All	kelp	55021	379088	434109	94.73	2.92	3.32
OA = 97.0736%	water	3062	12621571	12624633	5.27	97.08	96.68
KC = 0.2174	total	58083	13000659	13058742	100	100	100
MaxLike - Co-Oc Mean Cor	kelp	56570	340002	396572	97.4	2.62	3.04
OA = 97.3848%	water	1513	12660657	12662170	2.6	97.38	96.96
KC = 0.2430	total	58083	13000659	13058742	100	100	100
MaxLike - Co-Oc Mean Var	kelp	48751	162867	211618	83.93	1.25	1.62
OA = 98.6814%	water	9332	12837792	12847124	16.07	98.75	98.38
KC = 0.3570	total	58083	13000659	13058742	100	100	100
MaxLike - DR5 and Ent8	kelp	56927	420868	477795	98.01	3.24	3.66
OA = 96.7683%	water	1156	12579791	12580947	1.99	96.76	96.34
KC = .2062	total	58083	13000659	13058742	100	100	100
MaxLike - DR5 and Mean6 and Ent8	kelp	57428	342358	399786	98.87	2.63	3.06
OA = 97.3733%	water	655	12658301	12658956	1.13	97.37	96.94
KC = 0.2450	total	58083	13000659	13058742	100	100	100
MaxLike - DR5 and Var11	kelp	54803	221508	276311	94.35	1.7	2.12
OA = 98.2786%	water	3280	12779151	12782431	5.65	98.3	97.88
KC = 0.3228	total	58083	13000659	13058742	100	100	100
MaxLike - Pan 2 and Var11 and Cor17	kelp	53476	204366	257842	92.07	1.57	1.97
OA = 98.3997%	water	4607	12796293	12800900	7.93	98.43	98.03

KC = 0.3337	total	58083	13000659	13058742	100	100	100
MaxLike - Pan Data Var	kelp	55393	344335	399728	95.37	2.65	3.06
OA = 97.3426%	water	2690	12656324	12659014	4.63	97.35	96.94
KC = 0.2361	total	58083	13000659	13058742	100	100	100
MaxLike - Pan Var	kelp	49875	232861	282736	85.87	1.79	2.17
OA = 98.1540%	water	8208	12767798	12776006	14.13	98.21	97.83
KC = 0.2874	total	58083	13000659	13058742	100	100	100
MinDis - Pan Data Var	kelp	29159	45209	74368	50.2	0.35	0.57
OA = 99.4323%	water	28924	12955450	12984374	49.8	99.65	99.43
KC = 0.4375	total	58083	13000659	13058742	100	100	100
MinDis - Pan Var	kelp	28927	44303	73230	49.8	0.34	0.56
OA = 99.4375%	water	29156	12956356	12985512	50.2	99.66	99.44
KC = 0.4378	total	58083	13000659	13058742	100	100	100

Table 5. Classification Methods Confusion Matrices

These confusion matrices in Table 5 represent the different kinds of Classification Methods used and the associated texture features that were use in each. The regions-of-interest were used for the truth image kelp and water ROIs. Also, the user could just select an area that seems to be open water and kelp for the ROIs associated with each. The other confusion matrices are in order of the band threshold, and the results progress in a linear fashion, but the classification confusion matrices vary depending on the both the classification method and the texture features used. Overall, the classification methods performed well like the band thresholds, with overall accuracies of 96% to 99%. The Maximum Likelihood classification method was really good at detecting all of the kelp pixels no matter which texture features were used, but the amount of sunglint pixels increased and decreased depending on the texture features. The Maximum Likelihood classification method performed better than the Pan and Variance band threshold in eliminating the sunglint when detecting mostly all of the kelp. But the higher levels of the band thresholds have less sunglint pixels in them, which lead them to have higher overall accuracies than the Maximum Likelihood classification method.

The Binary Encoding classification method did a good job of reducing the sunglint pixels to 162780 but only correctly detected 76% of the kelp. The Mahalanobis

Distance classification method did an even better job reducing the sunglint pixels at 96256, but still only accurately detected 76% of the kelp. The Minimum Distance classification method did the best job of reducing the number of sunglint pixels to 45209 but only correctly detected 50% of the kelp pixels. With this big reduction in the sunglint pixels, the overall accuracy of this method was 99.4323% and is the highest of the classification methods.

5. Study Area Pan Image

		Ground Truth (Pixels)			Ground Truth (%)		
Gray Tone Value	feature	kelp NDVI	Water NDVI	total	kelp NDVI	Water NDVI	total
80	kelp	15943	602441	618384	99.97	89.19	89.44
OA = 12.8667%	water	4	73018	73022	0.03	10.81	10.56
KC = .0055	total	15947	675459	691406	100	100	100
85	kelp	15627	161636	177263	97.99	23.93	25.64
OA = 76.5758%	water	320	513823	514143	2.01	76.07	74.36
KC = .1247	total	15947	675459	691406	100	100	100
90	kelp	14536	42712	57248	91.15	6.32	8.28
OA = 93.6184%	water	1411	632747	634158	8.85	93.68	91.72
KC = 0.3746	total	15947	675459	691406	100	100	100
95	kelp	13499	20368	33867	84.65	3.02	4.9
OA = 96.7001%	water	2448	655091	657539	15.35	96.98	95.1
KC = 0.5271	total	15947	675459	691406	100	100	100
100	kelp	12455	11552	24007	78.1	1.71	3.47
OA = 97.8241%	water	3492	663907	667399	21.9	98.29	96.53
KC = 0.6127	total	15947	675459	691406	100	100	100
105	kelp	11454	6329	17783	71.83	0.94	2.57
OA = 98.4348%	water	4493	669130	673623	28.17	99.06	97.43
KC = .6712	total	15947	675459	691406	100	100	100
110	kelp	10433	3405	13838	65.42	0.5	2
OA = 98.7100%	water	5514	672054	677568	34.58	99.5	98
KC = 0.6940	total	15947	675459	691406	100	100	100
115	kelp	9477	1715	11192	59.43	0.25	1.62
OA = 98.8162%	water	6470	673744	680214	40.57	99.75	98.38
KC = 0.6926	total	15947	675459	691406	100	100	100
120	kelp	8535	869	9404	53.52	0.13	1.36
OA = 98.8023%	water	7412	674590	682002	46.48	99.87	98.64
KC = 0.6677	total	15947	675459	691406	100	100	100
125	kelp	7615	479	8094	47.75	0.07	1.17
OA = 98.7256%	water	8332	674980	683312	52.75	99.93	98.83
KC = 0.6277	total	15947	675459	691406	100	100	100

Table 6. Study Area PAN Image Confusion Matrices

These confusion matrices in Table 6 are the PAN thresholds for the study areas of Arroyo Burro and Mohawk. These confusion matrices get to the answer of how close the kelp can be detected for the kelp study by a simple threshold. The PAN thresholds perform well just as in the large PAN images, but there are some differences worth noting. The area has a lot less sunglint pixels as compared with the size of the kelp. The study area does not include the harbor, which was the main source of the sunglint pixels in the other images. This leads to a much higher Kappa Coefficient for the study area PAN band thresholds. The percentage of kelp pixels in the different images are 0.207% Large PAN, 0.447% Small PAN, and 2.36% Study Area PAN. The kelp and the sunglint pixels show the same rate of reduction across the band thresholds as seen in the other PAN images, as this is to be expected.

6. Study Area Variance Image

		Ground Truth (Pixels)			Ground Truth (%)		
Variance Value	feature	kelp NDVI	Water NDVI	total	kelp NDVI	Water NDVI	total
15	kelp	15764	25024	40788	98.85	3.7	5.9
OA = 96.3542%	water	183	650435	650618	1.15	96.3	94.1
KC = 0.5405	total	15947	675459	691406	100	100	100
30	kelp	15274	14417	29691	95.78	2.13	4.29
OA = 97.8175%	water	673	661042	661715	4.22	97.87	95.71
KC = 0.6591	total	15947	675459	691406	100	100	100
70	kelp	13235	7195	20430	82.99	1.07	2.95
OA = 98.5671%	water	2712	668264	670976	17.01	98.93	97.05
KC = 0.7204	total	15947	675459	691406	100	100	100
100	kelp	11665	5285	16950	73.15	0.78	2.45
OA = 98.6163%	water	4282	670174	674456	26.85	99.22	97.55
KC = .7021	total	15947	675459	691406	100	100	100
150	kelp	9324	3526	12850	58.47	0.52	1.86
OA = 98.5321%	water	6623	671933	678556	41.53	99.48	98.14
KC = 0.6402	total	15947	675459	691406	100	100	100
200	kelp	7300	2519	9819	45.78	0.37	1.42
OA = 98.3850%	water	8647	672940	681587	54.22	99.63	98.58
KC = 0.5589	total	15947	675459	691406	100	100	100
250	kelp	5754	1888	7642	36.08	0.28	1.11
OA = 98.2527%	water	10193	673571	683764	63.92	99.72	98.89
KC = 0.4801	total	15947	675459	691406	100	100	100

Table 7. Study Area Variance Confusion Matrices

These confusion matrices in Table 7 are the Variance thresholds for the study areas of Arroyo Burro and Mohawk. These confusion matrices as with the PAN thresholds help to the answer of how close the kelp can be detected for the kelp study. In this area, the lower Variance thresholds do even better than the other variance thresholds in detecting the kelp and eliminating the sunglint pixels. Also, as in the other case, the variance comes to a point where the Pan thresholds do better than the Variance thresholds.

7. Study Area Classification Image

		Ground Truth (Pixels)			Ground Truth (%)		
Classification	feature	kelp NDVI	Water NDVI	total	kelp NDVI	Water NDVI	total
BinEn - Pan Var	kelp	10699	5585	16284	67.09	0.83	2.36
OA = 98.4332%	water	5248	669874	675122	32.91	99.17	97.64
KC = .6559	total	15947	675459	691406	100	100	100
MahDis - Pan Var	kelp	11434	6322	17756	71.7	0.94	2.57
OA = 98.4329%	water	4513	669137	673650	28.3	99.06	97.43
KC = .6705	total	15947	675459	691406	100	100	100
MaxLike - Pan Var	kelp	13245	18032	31277	83.06	2.67	4.52
OA = 97.0012%	water	2702	657427	660129	16.94	97.33	95.48
KC = 0.5471	total	15947	675459	691406	100	100	100
MinDis - Pan Var	kelp	10063	3890	13953	63.1	0.58	2.02
OA = 98.5864%	water	5884	671569	677453	36.9	99.42	97.98
KC = 0.6659	total	15947	675459	691406	100	100	100

Table 8. Study Area Classification Methods Confusion Matrices

These confusion matrices in Table 8 are the Classification Methods for the study areas of Arroyo Burro and Mohawk. The four different types of classification Methods; Maximum Likelihood, Binary Encoding, Mahalanobis Distance, and Minimum Distance were used to classify the kelp and water. The Maximum Likelihood method did the best at detecting the highest percentage of kelp at 83%, but also had the highest amount of false detections at 18032 pixels, which are twice that of the number of kelp pixels. The other three, Binary Encoding, Mahalanobis Distance, and Minimum Distance, all detected less kelp in the 60%–70% range, but performed much better in the false detections in the 4000–6000 range. These allowed the overall detections of kelp to be very close to the actual number of kelp. Binary Encoding was 16284/15947, which equals 102.1%. Mahalanobis Distance was 17756/15947, which equals 111.3%.

Minimum Distance was 13953/15947, which equals 87.5%. These three classification methods all do a better job at eliminating the outliers of sunglint than Maximum Likelihood.

E. ROC CURVES

A receiver operating characteristic (ROC) is a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied. The ROC curve is represented by plotting the fraction of true positives vs. the fraction of false positives. A completely random guess would give a point along a diagonal line, the line of no-discrimination, from the left bottom (0,0) to the top right (1,1) corners. The higher the curve is relative to the line of no-discrimination, the better the method is at correctly guessing the pixels. (Fawcett, 2004) In the case of the intensity level of the PAN image, kelp had an extremely high ROC curve. Again, an important point to keep in mind is that the x-axis is affected by the large number of chances for false positives in the case of water and kelp. The low number of false positives leads to this extremely good ROC curve.

1. Large Pan Curve

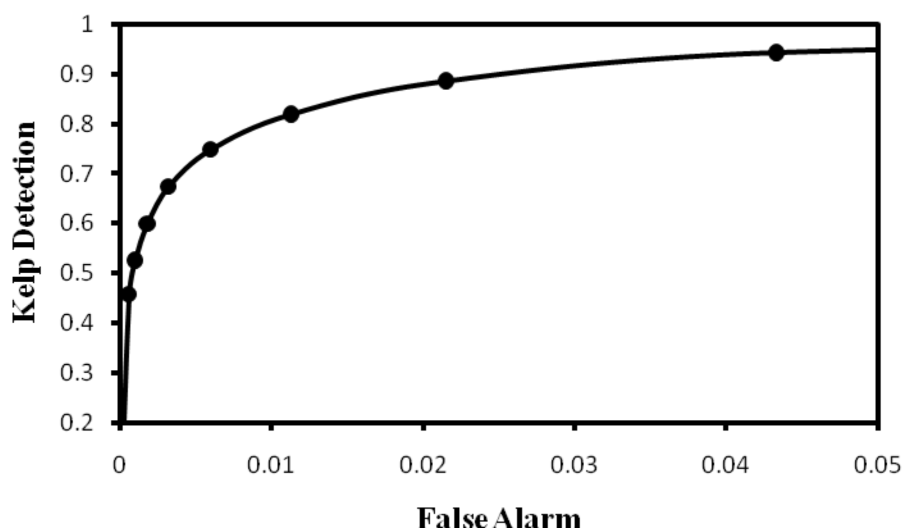


Figure 36. Zoomed Large PAN Image ROC curve

The plot in Figure 36 was of the large PAN image. With the large number of water pixels and the correct classification of them, the ROC curve is shifted far over to the left. This shows that the Pan threshold method is well above the line of no-discrimination, so it is a very good detector and discriminator of kelp and water.

2. Small Pan and Variance Curves

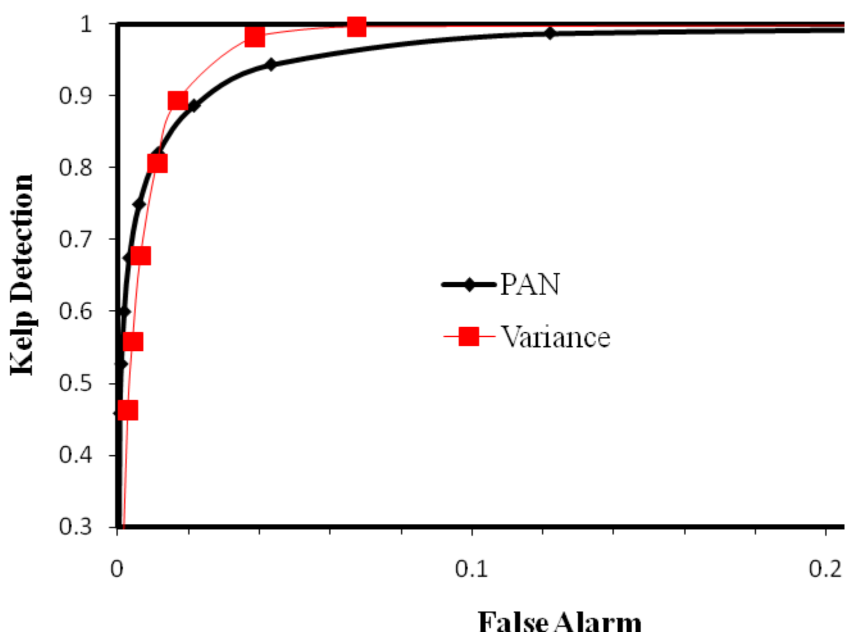


Figure 37. Small PAN Image and Variance ROC curve

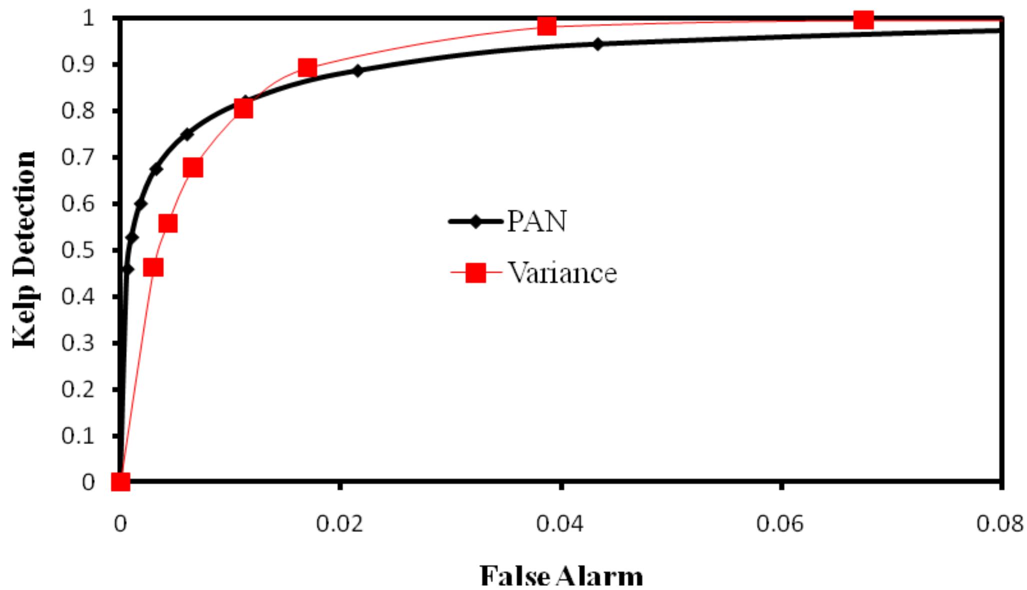


Figure 38. Zoomed Small PAN Image and Variance ROC curve

The plots in Figures 37 and 38 show the ROC curves of the small PAN image and the variance image. Again, as in the first set of plots, a second graph was needed to show the slope of the curve. Having both the PAN and Variance curves on the graph, the different characteristics can be compared. The variance curve does better in the beginning of the curve by having less false detections. Then, at a point, the small PAN curve takes over and becomes better. With both being this much over the no-discrimination line, it seems that these both would be useful in the classification methods.

3. Study Area Pan and Variance Curves for Study Area

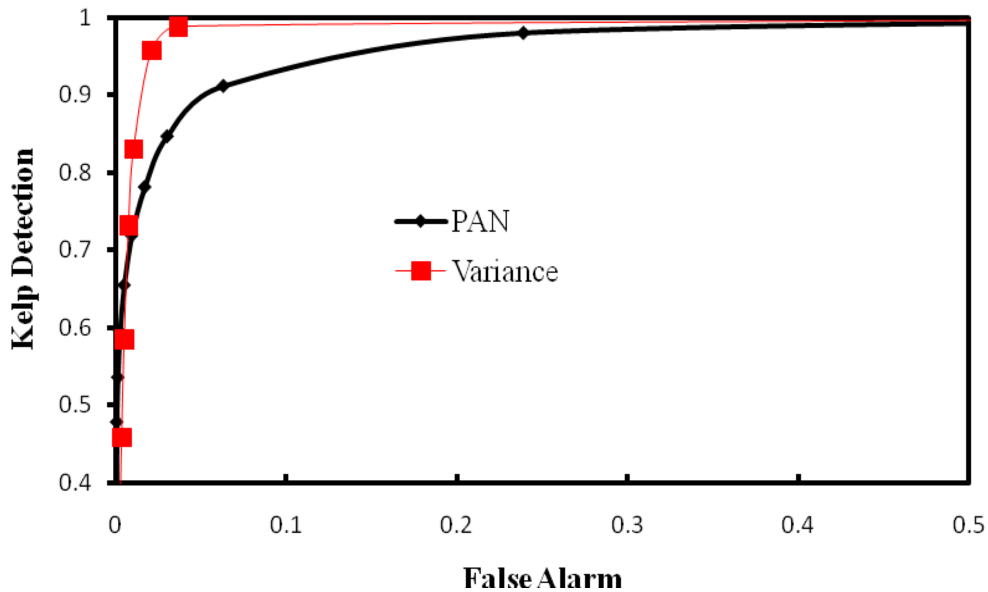


Figure 39. Study Area PAN Image and Variance ROC curve

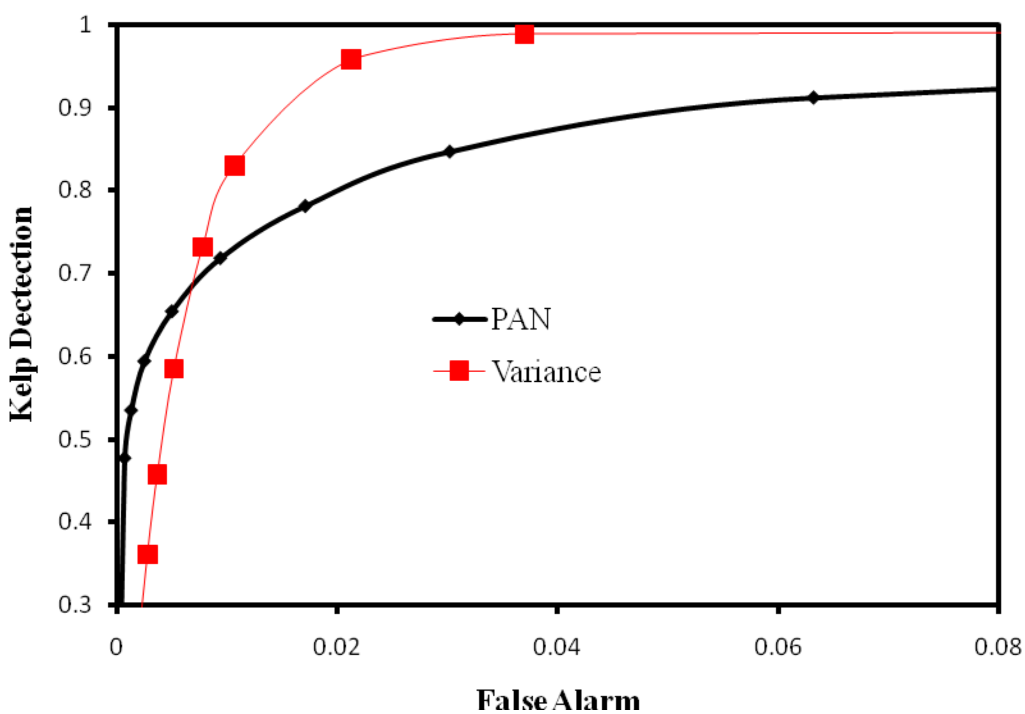


Figure 40. Zoomed Study Area PAN Image and Variance ROC curve

The plots in Figure 39 and 40 show the ROC curves of the small PAN image and the variance image. Again, as in the first set of plots, a second graph was needed to show the slope of the curve. The variance curve does even better in the beginning of the curve by having less false detections than the variance curves in Figure 40. This could be attributed to the sunglint pixels, but it does not show up in the PAN curve. The reason for the difference is not known. Then, at about the same point in Figure 40, the small PAN curve takes over and becomes better.

V. SUMMARY AND CONCLUSIONS

A. KELP DECTECTION

First, it has been shown that there is a strong linear relationship between the NDVI image and the PAN image. With this fact established, there is some noise that does not allow for a perfect classification of the kelp and water. Most all of the noise comes in the form of sunglint. Originally, sunglint was seen as one random occurrence that was caused by reflections on the water from the flicker of the tides. While this is true, it is not the only case. The white foam is associated with waves near the shore and the activity of boats near a harbor. This white foam proves harder to correctly classify. Other than the pan image, the variance texture feature seemed to be the best at discriminating the kelp from the water. Some of the other texture features did some level of discrimination, but did not add significantly to the overall accuracy of the selection of kelp or water.

The results in this study could be looked at in several different ways. The overall accuracies of even the Pan and variance band thresholds are very high, in the 98%-99% range, depending on which threshold is chosen. Now, if the user wants to just locate the kelp but noise is included, the lower band thresholds and the Maximum Likelihood Classification method do an excellent job. If the user wants to use only the closest number of pixels to get an overall area the same as the NDVI image, the Minimum Distance and the higher band thresholds do a much better job at this. The band threshold could vary from image to image, so this is not a desired method. The Minimum Distance only accounts for two-thirds of the true kelp but gives an overall area very close to the truth. More images would be needed to see how consistently close this method would be to the truth. The Mahalanobis Distance and Minimum Distance classification methods both performed as well as the best threshold values of the pan image, with overall accuracies of 98.5%. The pan threshold values are not a desired way to classify kelp because the value could vary from image to image. The optimal method to classify the kelp and water seems to be either Mahalanobis Distance or Minimum Distance classification methods using the pan and variance texture feature.

Overall, it proved difficult to eliminate all sunglint in the classification of kelp. Also, a small percentage of kelp can be lost in trying to eliminate sunglint. To answer the question, Can kelp be accurately classified by a panchromatic image? The answer is yes. A panchromatic image can correctly classify kelp and water, with an overall accuracy of 98% to 99%.

LIST OF REFERENCES

Alhaddad, B. (2008). Remote sensing flood hazards assessment using panchromatic satellite imagery. *ISPRS Congres* (pp. 881–886). Beijing: isprs.

Cavanaugh, K., Siegel, D. A., Kinlan, B. P., & Reed, D. C. (2007). Remote sensing of giant kelp forest cover and biomasssing SPOT multispectral imagery. *American Geophysical Union.*

Dayton, P. (1985). Ecology of kelp communities. *Annual Review of Ecology and Systematics*, 16: 215–245.

Fawcett, T. (2004). *ROC Graphs: Notes and Practical Considerations for Researchers.* Palo Alto: HP Laboratories.

Hall-Beyer, M. (2000). GLCM texture: A tutorial. *National Council on Geographic Information and Analysis Remote Sensing Core Curriculum*.

Haralick, R., Shanmugan, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*(6): 610–621.

Hayne, G. (1980). Radar altimeter mean return waveform from near-normal-incidence ocean surface scattering. *IEEE Transactions on Antenna and Propagation*, AP-28: 687–692.

Jackson, G., & Winant, C. (1983). Effect of a kelp forest on coastal currents. *Continental Shelf Report*, 2: 75–80.

Jones, C. J. (1997). Positive and negative effects of organisms as physical ecosystem engineers. *Ecology*, 1946–1957.

Lenihan, H. (2004, May 14). UCSB research explores kelp forest ecology. *Coastal View News*.

Mazer, A. S., Martin, M., Lee, M., & and Solomon, J. E. (1988). Image processing software for imaging spectrometry analysis. *Remote Sensing of Enviroment*, 24: 201–210.

Perry, C., & Lautenschlager, L. F. (1984). Functional equivalence of spectral vegetation indices. *Remote Sensing of Environment*.

Richards, J. (1999). *Remote Sensing Digital Image Analysis*. Berlin: Springer-Verlag.

Sala, E., Bourdouresque, C., & Harmelin-Vivien, M. (1998). Fishing, trophic cascades, and the structure of algal assemblages: evaluation of an old but untested paradigm. *Oikos*, 82: 425–439.

Shi, P., Zhang, Q., & Gong, P. (2003). Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing*, *24*(21): 4137–4160.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Col. Dave Madden
   Global Positioning Systems Wing
   El Segundo, California